DUKE MATHEMATICAL JOURNAL

EDITED BY

VOLUME 7
1940

DUKE UNIVERSITY PRESS
DURHAM, N. C.

CONTENTS

Volume 7, 1940

NECESSARY CONDITIONS IN GENERALIZED-CURVE PROBLEMS OF THE CALCULUS OF VARIATIONS

By E. J. McShane

In a preceding paper[1] we have developed the theory of generalized curves originated by L. C. Young.[2,3,4] For the problems of the calculus of variations in which a generalized curve is sought which minimizes an integral we have developed certain existence theorems. In a subsequent paper we shall obtain conditions under which the minimizing (generalized) curve can be shown to be an ordinary curve. But for this purpose we must develop the theory of the calculus of variations for generalized curves; and it is this development which forms the subject matter of the present paper. We shall develop necessary conditions for a minimum in problems of Bolza type.

Our extension of the theory must necessarily include, as a special case, the proof of the multiplier rule for rectifiable (ordinary) curves. Such a proof has already been given by Graves.[5] The non-parametric problem with generalized curves has been studied by L. C. Young;[6] however, Young considered only plane problems without side conditions, and we require a more extended theory. Moreover, we need a result not obtained by Young; it is important for us to know that for almost all t the partial derivatives F_{r^i} are constant over the set of vectors r carried by the minimizing curve C_0^* at $y_0(t)$. (Young obtained this result in a special case.[7]) Still further, in order to establish our existence theorems it is vital to know that the Weierstrass condition is satisfied along the minimizing curve, whether or not it is normal. This new form of the multiplier rule, including the Weierstrass condition, was first established in this note; the subsequent specialization to ordinary curves has already been published.[8]

Received November 28, 1939.

[1] E. J. McShane, *Generalized curves*, this Journal, vol. 6(1940), pp. 513–536; henceforth referred to by the letters GC.

[2] L. C. Young, *On approximation by polygons in the calculus of variations*, Proc. Royal Soc., (A), vol. 141(1933), pp. 325–341.

[3] L. C. Young, *Generalized curves and the existence of an attained absolute minimum in the calculus of variations*, Comptes Rendus de la Société des Sciences et des Lettres de Varsovie, Classe III, vol. 30(1937), pp. 212–234.

[4] L. C. Young, *Necessary conditions in the calculus of variations*, Acta Math., vol. 69(1938), pp. 229–258.

[5] L. M. Graves, *On the problem of Lagrange*, Amer. Jour. of Math., vol. 53(1931), pp. 547–554.

[6] Loc. cit. (footnote 4).

[7] Loc. cit. (footnote 4), pp. 248, 249.

[8] E. J. McShane, *On multipliers for Lagrange problems*, Amer. Jour. of Math., vol. 61(1939), pp. 809–819.

I am glad to express my indebtedness to Professor Graves for his careful reading of the manuscripts of these papers and for his helpful suggestions.

1. **Hypotheses.** Throughout this paper the notation and terminology of GC will be retained, with a single exception; the repetition of any affix (Latin or Greek) in a term will indicate that that term is to be summed over all the values of the repeated affix.

In this section we collect some definitions and some hypotheses which will be assumed satisfied throughout the remainder of this paper.

(1.1) *R is an open set of points in (y, r)-space (i.e., in $(y^1, \cdots, y^\nu, r^1, \cdots, r^\nu)$-space) such that if (y, r) is in R, so is (y, kr) for all positive numbers k.*

(1.2) *The functions $f(y, r)$, $\varphi^\alpha(y, r)$ $(\alpha = 1, \cdots, m < \nu - 1)$ are defined on R and are positively homogeneous of degree 1 in r. Also, they are of class C^κ, where κ is at least 1; that is, all their partial derivatives of order $\leqq \kappa$ are defined and continuous for all (y, r) in R with $|r| \neq 0$.*

(1.3) Definition. A point (y, r) is admissible if it belongs to R, has $|r| \neq 0$, and satisfies the equations

$$\varphi^\alpha(y, r) = 0 \qquad (\alpha = 1, \cdots, m),$$

and the matrix

$$\| \varphi^\alpha_{r^i}(y, r) \|$$

has rank m.

(1.4) Definition. If $C^*: [y(t), \mathfrak{M}[t; \Phi], M]$ is a generalized curve, we define $R(C^*)$ to be the closure of the aggregate of points $(y(t), r)$ with t in M and r carried at $y(t)$. (This actually may depend on the representation and not on the curve alone, since rejecting a single point from M may diminish $R(C^*)$; but this will cause no trouble.)

Our definition of admissibility of a curve may seem rather stringent. However, for the minimizing curve it is the analogue of a condition commonly assumed, and there is certainly no loss of generality in restricting the class of comparison curves.

(1.5) Definition. A generalized curve C^* is admissible if every element (y, r) in $R(C^*)$ is admissible.

If C^* is admissible, the function $f(y, r)$ is continuous on the closed set $R(C^*)$, and so there is a function $f_0(y, r)$ which is defined and continuous for all y and all r and which coincides with $f(y, r)$ on $R(C^*)$. Then by §3 of GC the integral

$$\int_a^b \mathfrak{M}[t; f_0(y(t), r)]\, dt$$

exists, where $[a, b]$ is $\bar{M}$. But for all t in M the integrand depends only on the values of $f_0(y(t), r)$ on the set of vectors r carried at $y(t)$, and this is uniquely determined by the function $f(y, r)$. Thus the integral above is determined

uniquely by $f(y, r)$ and the curve C^*, and we can designate it by either of the symbols

$$\mathcal{F}(C^*) = \int_a^b \mathfrak{M}[t; f(y(t), r)]\, dt.$$

Our next standing hypothesis is

(1.6) *P is an open set in the 2ν-dimensional space of points $(y_1^1, \cdots, y_1^\nu, y_2^1, \cdots, y_2^\nu)$, and the functions $g(y_1, y_2)$ and $\psi^\mu(y_1, y_2)$ $(\mu = 1, \cdots, p \leqq 2\nu)$ are defined and of class C′ on P.*

(1.7) *C_0^* is an admissible generalized curve, given in standard representation by the formula $[y_0(t), \mathfrak{M}_0[t; \Phi], M_0]$, whose end points $(y_0(0), y_0(1))$ determine a point in P and satisfy the equations*

$$\psi^\mu(y_0(0), y_0(1)) = 0 \qquad (\mu = 1, \cdots, p);$$

and it gives the functional

$$J(C^*) \equiv g(y(a), y(b)) + \int_a^b \mathfrak{M}[t; f(y(t), r)]\, dt$$

(the notation for C^* being as in (1.4)) *a strong relative minimum*[9] *in the class of all admissible generalized curves whose end points $(y(a), y(b))$ determine a point in P and satisfy the equations $\psi^\mu(y(a), y(b)) = 0$.*

(1.8) *The matrix*

$$\| \psi^\mu_{y_1^i}(y_0(0), y_0(1)), \psi^\mu_{y_2^i}(y_0(0), y_0(1)) \| \qquad (\mu = 1, \cdots, p;\ i = 1, \cdots, \nu)$$

has rank p.

Our next hypothesis represents an unnecessary restriction on the generality of our theorems; the theorems can be established without it. But the loss of generality is not very great, and the gain in convenience is so considerable that it seems worth while to adopt the restriction. For each admissible (y, r) the matrix in (1.4) has rank m, so we can adjoin $\nu - m$ rows in such a way that the resulting square matrix is non-singular. Our hypothesis is to the effect that this can be done once and for all on $R(C_0^*)$. That is,

(1.9) *there are functions $\varphi^{m+1}(y, r), \cdots, \varphi^\nu(y, r)$, defined and of class C″ on R, such that the square matrix*

$$\| \varphi^i_{r^j}(y, r) \| \qquad (i, j = 1, \cdots, \nu)$$

is non-singular for all sets (y, r) in $R(C_0^*)$.

The following statement is an immediate consequence of (1.9) and the closure of $R(C_0^*)$.

(1.10) Corollary. *There is a neighborhood U of $R(C_0^*)$ such that $\bar{U}$ is contained in R and the matrix in (1.9) is non-singular for all (y, r) in $\bar{U}$.*

[9] The concept of the strong relative minimum is the same as for ordinary curves; there is a positive ϵ such that $J(C^*) \geqq J(C_0^*)$ if C^* satisfies the conditions specified in (1.7) and the Fréchet distance between the tracks of C^* and C_0^* is less than ϵ.

2. A lemma on multipliers. In accordance with custom, for an arbitrary set of numbers $\lambda^0, \lambda^1, \cdots, \lambda^\nu$ we define

$$F(y, r, \lambda) = \lambda^0 f(y, r) + \lambda^i \varphi^i(y, r). \tag{2.1}$$

The partial derivatives F_{y^i}, F_{r^i} are computed from this, the λ^i being treated as independent variables.

LEMMA 2.1. *For every set of constants λ^0, $c_1, \cdots, c_\nu$ there exist functions $\lambda^i(t, r)$ $(i = 1, \cdots, \nu)$, defined and continuous on $\bar{U}$ and of class $C^{\kappa-1}$ in r for each fixed t, such that the equations*

$$F_{r^i}(y_0(t), r, \lambda(t, r)) = c_i + \int_0^t \mathfrak{M}_0[t; F_{y^i}(y_0(t), r, \lambda(t, r))]\,dt \tag{2.2}$$

hold for all (t, r) such that t is in $[0, 1]$ *and $(y_0(t), r)$ is in $\bar{U}$.*

In equation (2.2) and all later expressions of this type, we understand $\lambda^0(t, r)$ to be the constant λ^0.

By (1.10), the matrix in hypothesis (1.9) has an inverse $\| \psi_k^j(y, r) \|$ such that each $\psi_k^j(y, r)$ is defined and of class $C^{\kappa-1}$ on $\bar{U}$. We shall now show that it is possible to determine absolutely continuous functions $h^j(t)$ $(0 \leqq t \leqq 1$; $j = 1, \cdots, \nu)$ in such a way that the functions

$$\lambda^i(t, r) = \{-\lambda^0 f_{r^j}(y_0(t), r) + h^j(t)\}\psi_i^j(y_0(t), r) \tag{2.3}$$

have the desired properties. By substitution we find

$$F_{r^i}(y_0(t), r, \lambda(t, r)) = h^i(t), \tag{2.4}$$

while equation (2.2) takes the form

$$h^i(t) = c_i + \int_0^t \mathfrak{M}_0[t; \lambda^0\{f_{y^i} - f_{r^j}\psi_k^j \varphi_{y^i}^k\} + h^j(t)\psi_k^j \varphi_{y^i}^k]\,dt, \tag{2.5}$$

the arguments of the functions in the integrand being $(y_0(t), r)$.

The integrands on the right are measurable and uniformly bounded in t for fixed h, and are uniformly Lipschitzian in h for fixed t. Hence[10] these equations have unique solutions $h^j(t)$, and the lemma is established.

Remark. If desired, we may suppose that the $\lambda^i(t, r)$ are defined and continuous and of class $C^{\kappa-1}$ in r for all t and r. The possibility of this extension will be established in §18.

3. Embedding a curve in a family. The generalized curve C_0^* will be said to be *embedded* in a family of admissible curves

$$C_b^* : [y(t, b), \mathfrak{M}_b[t; \Phi], M_0],$$

where b is $(b_1, \cdots, b_{q^*})$, if the following conditions are satisfied.

[10] Carathéodory, *Vorlesungen über reelle Funktionen*, p. 674.

(3.1) *For all b in a neighborhood of* $(0, \cdots, 0)$ *the curve C_b^* is an admissible generalized curve.*
(3.2) *The functions $y(t, b)$ and their partial derivatives with respect to the b_q are defined and continuous for all t in* $[0, 1]$ *and all b near* $(0, \cdots, 0)$, *and they satisfy the equation* $y(t, 0) = y_0(t)$.
(3.3) *The means $\mathfrak{M}_b[t; \Phi]$ are defined by an equation*

$$\mathfrak{M}_b[t; \Phi] = \mathfrak{M}_0[t; \Phi^*], \tag{3.3a}$$

where

$$\Phi^*(r) = \Phi(\bar{r}(t, r, b)); \tag{3.3b}$$

the functions $\bar{r}^i(t, r, b)$ are defined and continuous together with their partial derivatives with respect to the b_q for all b near zero and all (t, r) such that $(y_0(t), r)$ lies in a certain neighborhood U_1 of $R(C_0^*)$, and for $b = (0, \cdots, 0)$ *they satisfy*

$$\bar{r}^i(t, r, 0) = r^i. \tag{3.3c}$$

If desired, in (3.3) we could require that the $\bar{r}^i(t, r, b)$ be defined and continuous, together with their partials with respect to the b_q, for all t, r, and b. This would not change the content of the definition, as is shown by Lemma 18.1. We shall use the notation

$$\begin{aligned} \eta_q^i(t) &= \frac{\partial}{\partial b_q} y^i(t, b)\bigg|_{b=0}, \\ \rho_q^i(t, r) &= \frac{\partial}{\partial b_q} \bar{r}^i(t, r, b)\bigg|_{b=0}. \end{aligned} \tag{3.4}$$

From (3.3) we see readily that
(3.5) *the vector r_0 is carried by C_0^* at $y_0(t)$ if and only if $\bar{r}(t, r_0, b)$ is carried by C_b^* at $y(t, b)$.*

From (3.5) it follows that for b near $(0, \cdots, 0)$ the set $R(C_b^*)$ is in an arbitrarily small neighborhood of $R(C_0^*)$, hence in particular is in R. Thus in the presence of (3.2) and (3.3), hypothesis (3.1) is equivalent to demanding that for each t in M_0 and each b near $(0, \cdots, 0)$ the functions $\bar{r}(t, r, b)$ shall map the vectors carried by C_0^* at $y_0(t)$ into the set of solutions of the equations $\varphi^\alpha(y(t, b), r) = 0$.

In particular, if t is in M_0 and r_0 is a vector carried by C_0^* at $y_0(t_0)$, the equations

$$\varphi^\alpha(y(t, b), \bar{r}(t, r_0, b)) = 0 \tag{3.6}$$

must hold for all b near $(0, \cdots, 0)$. If we differentiate and set $b = 0$, this yields

$$\varphi^\alpha_{y^i}(y_0(t), r_0)\eta_q^i(t) + \varphi^\alpha_{r^i}(y_0(t), r_0)\rho_q^i(t, r_0) = 0 \qquad (\alpha = 1, \cdots, m; t \text{ in } M_0). \tag{3.7}$$

We now adopt a well-known technique; corresponding to each $[\eta(t), \rho(t, r)]$ we introduce the functions $\zeta^i(t, r)$ defined by the equations

$$\zeta^i(t, r) = \varphi^i_{y^j}(y_0(t), r)\eta^j(t) + \varphi^i_{r^j}(y_0(t), r)\rho^j(t, r) \qquad (i, j = 1, \cdots, \nu). \tag{3.8}$$

From the preceding, we see that, if C_0^* is embedded in a family C_b^*,
(3.9) *the functions* $\zeta_q^i(t, r)$ *are defined and continuous for all* t *in* [0, 1] *and all* r *such that* $(y_0(t), r)$ *is in* $\bar{U}$, *and they satisfy the equations*

$$\zeta_q^\alpha(t, r_0) = 0 \qquad (\alpha = 1, \cdots, m;\ q = 1, \cdots, q^*)$$

for all t *in* M_0 *and all* r_0 *carried by* C_0^* *at* $y_0(t)$.

4. **The embedding theorem.** We now proceed to establish the embedding theorem.

THEOREM 4.1. *Let* $\zeta_q^i(t, r)$ $(q = 1, \cdots, q^*;\ i = 1, \cdots, \nu)$ *be a set of functions satisfying* (3.9). *Let* t_0 *be a number in the interval* [0, 1]. *Let* $Y^i(b)$ $(i = 1, \cdots, \nu)$ *be continuously differentiable functions of the* q^* *parameters* b_q, *defined for* b *near* $(0, \cdots, 0)$, *and such that* $Y(0) = y_0(t_0)$. *Then* C_0^* *can be embedded in a* q^**-parametered family of admissible curves* C_b^* *for which*

$$y^i(t_0, b) = Y^i(b), \tag{4.1}$$

and such that on the set $\bar{U}$ *the equations*

$$\varphi_{y^j}^i(y_0(t), r)\eta_q^j + \varphi_{r^j}^i(y_0(t), r)\rho_q^j(t, r) = \zeta_q^i(t, r) \qquad (i = 1, \cdots, \nu;\ q = 1, \cdots, q^*) \tag{4.2}$$

are satisfied.

The equations

$$\varphi^i(y, \bar{r}) = \varphi^i(y_0(t), r) + b_q\zeta_q^i(t, r) \tag{4.3}$$

have the initial solutions $\bar{r} = r$ on the set
(4.4) $0 \leqq t \leqq 1$, $y = y_0(t)$, r *such that* $(y_0(t), r) \in \bar{U}$, $b = 0$.
On this set the Jacobian of the left members of equations (4.3) with respect to the $\bar{r}^i$ is the determinant of the non-singular matrix in (1.9), and is therefore not zero. Hence[11] equations (4.3) have unique solutions $\bar{r}^i = \bar{r}^i(t, y, r, b)$ defined and continuous on a closed neighborhood $\bar{V}$ of the set (4.4) and reducing to the initial solution $\bar{r} = r$ on that set. These functions are continuously differentiable as functions of y and b for fixed t and r. We suppose them extended so as to be defined and continuous for all t, y, r and b.

Consider now the "differential equations"

$$y^i = Y^i(b) + \int_{t_0}^{t} \mathfrak{M}_0[t; \bar{r}^i(t, y, r, b)]\, dt. \tag{4.5}$$

For fixed y and b, the integrands in the right members are uniformly bounded functions of t, since the $\bar{r}^i$ are bounded; and they are measurable, by Lemma 3.1 of GC. We wish to show that for fixed t they are continuously differentiable functions of y and b. Suppose that $\psi(c, r)$ is a function defined and continuous

[11] G. A. Bliss, *Fundamental Existence Theorems*, The Princeton Colloquium (Amer. Math. Soc.), New York, 1913, p. 20.

with its partial derivative ψ_c for $c_1 < c < c_2$ and all r. If c is in (c_1, c_2) and ϵ near zero, we have by the linearity of $\mathfrak{M}_0$

$$\begin{aligned}\epsilon^{-1}\{\mathfrak{M}_0[t; \psi(c+\epsilon, r)] - \mathfrak{M}_0[t; \psi(c, r)]\} &= \mathfrak{M}_0[t; \epsilon^{-1}(\psi(c+\epsilon, r) - \psi(c, r))] \\ &= \mathfrak{M}_0\left[t; \epsilon^{-1}\int_0^\epsilon \psi_c(c+\tau, r)\, d\tau\right].\end{aligned} \tag{4.6}$$

In computing this last mean we need consider only values of r carried at $y_0(t)$, since $\mathfrak{M}_0$ is independent of other values. For such r we obtain, by the continuity of ψ_c, $\lim \epsilon^{-1}\int_0^\epsilon \psi_c(c+\tau, r)\, d\tau = \psi_c(c, r)$, uniformly on the set of r carried. Hence in (4.6) we may let ϵ approach zero and obtain

$$\frac{d}{dc}\mathfrak{M}_0[t; \psi(c, r)] = \mathfrak{M}_0[t; \psi_c(c, r)]. \tag{4.7}$$

Applying this in (4.5) we see that the integrands on the right are continuously differentiable functions of y and b for each fixed t. It follows[12] that the equations (4.5) have solutions $y^i = y^i(t, b)$ which are continuous with the partial derivatives

$$\frac{\partial}{\partial b_q} y^i(t, b)$$

for all t in $[0, 1]$ and all b near $(0, \cdots, 0)$. In particular, if we set $b = 0$, the functions $y_0^i(t)$ satisfy (4.5), and since the solution is unique, we have

$$y^i(t, 0) = y_0^i(t). \tag{4.8}$$

We now define $\bar{r}^i(t, r, b)$ to be the function $\bar{r}^i(t, y(t, b), r, b)$. Substituting in (4.5), we find

$$y^i(t, b) = Y^i(b) + \int_{t_0}^{t} \mathfrak{M}_0[t; \bar{r}^i(t, r, b)]\, dt. \tag{4.9}$$

If we define $\mathfrak{M}_b$ as in (3.3), the curves

$$C_b^*\colon\ [y(t, b); \mathfrak{M}_b[t; \Phi(r)], M_0]$$

satisfy (2.8c) of GC, as follows from (4.9), and satisfy (2.8d), by Lemma 3.1 of GC. The others of conditions (2.8) of GC are easily verified, so that C_b^* is a generalized curve.

Equations (4.3) can now be written in the form

$$\varphi^i(y(t, b), \bar{r}(t, r, b)) \equiv \varphi^i(y_0(t), r) + b_q\zeta_q^i(t, r), \tag{4.10}$$

valid for all (t, r) such that $(y_0(t), r)$ is in $\bar{U}$ and all b near $(0, \cdots, 0)$. By (3.5), a vector r is carried by C_0^* at $y_0(t)$ if and only if $\bar{r}(t, r, b)$ is carried by C_b^* at $y(t, b)$. But for such vectors the last term in the right member of equa-

[12] Carathéodory, op. cit. (footnote 10), pp. 672, 682.

tion (4.10) vanishes for $i = 1, \cdots, m$ by (3.9), and the first term vanishes for $i = 1, \cdots, m$ because C_0^* is admissible. Hence

$$\varphi^\alpha(y(t, b), \bar{r}) = 0 \qquad (\alpha = 1, \cdots, m)$$

for all vectors $\bar{r}$ carried by C_b^* at $y(t, b)$, and C_b^* is an admissible curve. We have thus verified conditions (3.1), (3.2) and (3.3), so that C_0^* is embedded in the q^*-parametered family C_b^* of admissible curves. Differentiating both members of (4.10) with respect to b_p and setting $b = 0$ yield the equations (4.2). Equation (4.1) follows at once from (4.9), and the theorem is established.

As an immediate corollary to Theorem 4.1 we have

THEOREM 4.2. *Let $\zeta_q^i(t, r)$ be a set of functions satisfying* (3.9), *and* l_q^i $(i = 1, \cdots, \nu;\ q = 1, \cdots, q^*)$ *a set of numbers. Then C_0^* can be embedded in a q^*-parametered family of admissible curves C_b^* such that* (4.2) *holds and*

$$\eta_q^i(1) = l_q^i. \tag{4.11}$$

We need only apply Theorem 4.1 with $t_0 = 1$ and

$$Y^i(b) = b_q l_q^i + y_0^i(1).$$

Another corollary is useful.

COROLLARY. *Under the hypotheses of Theorem* 4.1 *or* 4.2, *the functions $\eta_q(t)$ and $\rho_q(t, r)$ are uniquely determined.*

From (4.9) and (4.7) we obtain, on differentiating with respect to b_q and setting $b_1 = \cdots = b_{q^*} = 0$,

$$\eta_q^i(t) = Y_{b_q}^i(0) + \int_{t_0}^t \mathfrak{M}_0[t; \rho_q^i(t, r)]\, dt. \tag{4.12}$$

In (4.2) we multiply by the inverse $\psi_i^k(y_0(t), r)$ of the matrix $\varphi_{r^j}^i$, obtaining

$$\rho_q^k(t, r) = \psi_i^k(y_0(t), r)\{\zeta_q^i(t, r) - \varphi_{y^j}^i(y_0(t), r)\eta_q^j(t)\}. \tag{4.13}$$

This, with (4.12), yields

$$\eta_q^k(t) = Y_b^k(0) + \int_{t_0}^t \mathfrak{M}_0[t; \psi_i^k \zeta_q^i]\, dt - \int_{t_0}^t \eta_q^j(t)\mathfrak{M}_0[t; \psi_i^k \varphi_{y^j}^i]\, dt.$$

This equation has a unique solution,[13] so that $\eta_q(t)$ is uniquely determined. By (4.13), $\rho_q(t, r)$ is also uniquely determined.

5. Computation of the first variation. A set of functions $\zeta^i(t, r)$ satisfying (3.9) and a set of numbers $\eta^i(1)$ (such a system will henceforth be called an *admissible weak variation* $[\zeta(t, r), \eta(1)]$) determine, though not uniquely, a family of admissible curves C_b^* in which C_0^* is embedded. We wish to compute the first variation

$$\left.\frac{d}{db} J(C^*)\right|_{b=0}$$

[13] Carathéodory, loc. cit. (footnote 10).

of our functional (cf. (1.7))

$$
\begin{aligned}
J(C_b^*) &= g(y(0, b), y(1, b)) + \mathcal{F}(C_b^*),\\
\mathcal{F}(C_b^*) &= \int_0^1 \mathfrak{M}_0[t; f(y(t, b), \bar{r}(t, r, b))]\, dt.
\end{aligned}
\tag{5.1}
$$

The derivative of the first term in $J(C^*)$ is easily computed; it is

$$
\begin{aligned}
\left.\frac{dg}{db}\right|_{b=0} &= g_{y_1^i}(y_0(0), y_0(1))\eta^i(0) + g_{y_2^i}(y_0(0), y_0(1))\eta^i(1),\\
&= G(\eta(0), \eta(1)),
\end{aligned}
\tag{5.2}
$$

where G is simply an abbreviation for the sum preceding it in (5.2). It remains to compute the derivative of $\mathcal{F}(C_b^*)$.

In order to put this computation in a form useful for the next section, we define

$$
u(b) = \int_\alpha^1 \mathfrak{M}_0[t; f(y(t, b), \bar{r}(t, r, b))]\, dt,
$$

where α is an arbitrary fixed number in the interval [0, 1]. Thus if $\alpha = 0$ the function $u(b)$ is $\mathcal{F}(C_b^*)$. Using equation (4.7), we obtain

$$
\begin{aligned}
u'(b) &= \frac{d}{db}\int_\alpha^1 \mathfrak{M}_0[t; f(y(t, b), \bar{r}(t, r, b))]\, dt\\
&= \int_\alpha^1 \mathfrak{M}_0[t; f_{y^i}(y(t, b), \bar{r})y_b^i(t, b) + f_{r^i}(y(t, b), \bar{r})\bar{r}_b^i(t, r, b)]\, dt,
\end{aligned}
\tag{5.3}
$$

so that the derivative on the left is continuous near $b = 0$. (Evidently the same process would show that if there are q^* variations $[\zeta_q(t, r), \eta_q(1)]$ the partial derivative of $u(b)$ with respect to each b_q is continuous near $b = 0$.) Setting $b = 0$ yields

$$
u'(0) = \int_\alpha^1 \mathfrak{M}_0[t; f_{y^i}(y_0(t), r)\eta^i(t) + f_{r^i}(y_0(t), r)\,\rho^i(t, r)]\, dt.
\tag{5.4}
$$

By equation (4.2) with $q^* = 1$, we have

$$
\int_\alpha^1 \mathfrak{M}_0[t; \lambda^j(t, r)\{\varphi^j_{y^i}(y_0(t), r)\eta^i(t) + \varphi^j_{r^i}(y(t), r)\rho^i(t, r) - \zeta^j(t, r)\}]\, dt = 0.
\tag{5.5}
$$

If we multiply both members of (5.4) by λ^0 and use (5.5), we obtain, with the notation (2.1),

$$
\lambda^0 u'(0) = \int_\alpha^1 \mathfrak{M}_0[t; F_{y^i}(y_0(t), r, \lambda)\eta^i(t) + F_{r^i}(y_0(t), r, \lambda)\rho^i(t, r) - \lambda^j\zeta^j(t, r)]\, dt.
\tag{5.6}
$$

We suppose that for the arbitrary constants λ^0, c_i the functions $\lambda^i(t, r)$ have been chosen so as to satisfy equation (2.2). By (2.2) and (2.4) the function $h^i(t)$ satisfies the equation

$$h^i(t) \equiv c_i + \int_0^t \mathfrak{M}_0[t; F_{y^i}(y_0(t), r, \lambda)]\, dt. \tag{5.7}$$

Hence equation (5.6) can be written in the form

$$\lambda^0 u'(0) = \int_\alpha^1 \{\eta^i(t)\dot{h}^i(t) + h^i(t)\mathfrak{M}_0[t; \rho^i(t, r)] - \mathfrak{M}_0[t; \lambda^j(t, r)\zeta^j(t, r)]\}\, dt. \tag{5.8}$$

By (4.12), for almost all t the sum of the first two terms in the integrand is the derivative of the absolutely continuous function $\eta^i h^i$, so that

$$\lambda^0 u'(0) = \eta^i(t)h^i(t)\Big|_\alpha^1 - \int_\alpha^1 \mathfrak{M}_0[t; \lambda^j(t, r)\zeta^j(t, r)]\, dt. \tag{5.9}$$

In particular, if $\alpha = 0$ this equation and (5.2) yield

$$\begin{aligned}\lambda^0 \frac{d}{db} J(C_b^*)\Big|_{b=0} \\ = \lambda^0 G(\eta(0), \eta(1)) + \eta^i(t)h^i(t)\Big|_0^1 - \int_0^1 \mathfrak{M}_0[t; \lambda^j(t, r)\zeta^j(t, r)]\, dt.\end{aligned} \tag{5.10}$$

6. **Strong variations.** By Theorem 4.2 each set of admissible weak variations $[\zeta_q(t, r), \eta_q(1)]$ defines a family C_b^* satisfying equations (4.2) and (4.12). We now turn our attention to strong variations. By an *admissible strong variation* we shall mean a symbol of the form $[t_0, r_0]$, where $0 \leqq t \leqq 1$ and $(y_0(t_0), r_0)$ is admissible.

Suppose that we start with the family C_b^* of Theorem 4.2. Let $[t_0, r_0]$ be an admissible strong variation. We adjoin $\nu - m$ analytic functions $\bar{\varphi}^\beta(r)$ ($\beta = m + 1, \cdots, \nu$) such that the matrix

$$\left\| \begin{matrix} \varphi^\alpha_{r^i}(y, r) \\ \bar{\varphi}^\beta_{r^i}(r) \end{matrix} \right\| \tag{6.1}$$

is non-singular at $(y_0(t_0), r_0)$. By the usual methods we can then show that for all b near 0 the equations

$$\varphi^\alpha(y, y') = 0, \qquad \bar{\varphi}^\beta(y') - \bar{\varphi}^\beta(r_0) = 0 \tag{6.2}$$

have a unique solution $y = Y(s, b)$ (the prime denoting differentiation with respect to s) which satisfies the initial conditions

$$Y^i(0, b) = y^i(t_0, b). \tag{6.3}$$

In particular, at $s = b = 0$ we have

$$Y^i(0, 0) = y_0^i(t_0);$$

so by the non-singularity of (6.1) we find that

$$Y^{i\prime}(0, 0) = r_0^i. \tag{6.4}$$

We now apply Theorem 4.1 with the following notation. The parameters b of the preceding section are fixed at a value near zero; the parameter b of Theorem 4.1 will be designated by τ. For $Y(b)$ we choose $Y(\tau, b)$, where $Y(s, b)$ is defined by (6.2) and (6.3). In Theorem 4.1 we choose $\zeta(t, r) \equiv 0$. Then by that theorem there is a family of admissible curves

$$\bar{C}^*_{b,\tau}\colon [\bar{y}(t, b, \tau), \overline{\mathfrak{M}}_{b,\tau}[t; \Phi(r)], M]$$

embedding C^* and having tracks which satisfy the equation

$$\bar{y}^i(t_0, b, \tau) = Y^i(\tau, b). \tag{6.5}$$

Now for all *non-negative* τ near zero we define a generalized curve $C^*_{b,\tau}$ consisting of the following three arcs:

$$\begin{aligned}
&\text{First arc: } y^i = y^i(t, b),\ 0 \leqq t \leqq t_0,\\
&\quad \mathfrak{M}_{b,\tau}[t; \Phi(r)] \equiv \mathfrak{M}_b[t; \Phi(r)].\\
&\text{Second arc: } y^i = Y^i(s, b),\ 0 \leqq s \leqq \tau,\\
&\quad \mathfrak{M}_{b,\tau}[s; \Phi(r)] = \Phi(Y'(s, b)).\\
&\text{Third arc: } y^i = \bar{y}^i(t, b, \tau),\ t_0 \leqq t \leqq 1,\\
&\quad \mathfrak{M}_{b,\tau}[t; \Phi(r)] = \overline{\mathfrak{M}}_{b,\tau}[t; \Phi(r)].
\end{aligned} \tag{6.6}$$

From (6.2) and (6.5) we see that the track is continuous, and it is easy to verify that the conditions (2.8) of GC are all satisfied. (It would be easy to give a single parametrization of the curve, instead of the piecemeal representation here adopted, but this is of no importance.)

For the value of $\mathcal{F}(C^*_{b,\tau})$ we have

$$\begin{aligned}
\mathcal{F}(C^*_{b,\tau}) = \int_0^{t_0} \mathfrak{M}_b[t; f(y(t, b), r)]\,dt &+ \int_0^{\tau} f(y(s, b), Y'(s, b))\,ds\\
&+ \int_{t_0}^1 \overline{\mathfrak{M}}_{b,\tau}[t; f(\bar{y}(t, b, \tau), r)]\,dt.
\end{aligned} \tag{6.7}$$

The second term on the right is readily seen to be of class C′ in τ and b; so are the other two terms, as is shown by an argument like that at the beginning of §5. We now differentiate with respect to τ and set $b = \tau = 0$. The derivative of the last term on the right is calculated by (5.9), with $\alpha = t_0$. We use (6.4) and (6.2), and recall that $\zeta(t, r) = 0$; also, from (6.5) we compute $\bar{\eta}(t_0) = Y'(0, 0) = r_0$. The result is

$$\lambda^0 \frac{\partial}{\partial \tau} \mathcal{F}(C^*_{b,\tau})\bigg|_{b=\tau=0} = h^i(1)\,\bar{\eta}^i(1) + \lambda^0 f(y_0(t_0), r_0) - r_0^i h^i(t_0). \tag{6.8}$$

If we combine this with equation (5.2), recalling that the initial point of $C^*_{b,\tau}$ is independent of τ, we obtain

$$\lambda^0 \frac{\partial}{\partial \tau} J(C^*_{b,\tau})\Big|_{b=\tau=0} = \lambda^0 G(0, \bar{\eta}(1)) + h^i(1)\bar{\eta}^i(1) + \{\lambda^0 f(y_0(t_0), r_0) - r_0^i h^i(t_0)\}. \tag{6.9}$$

7. **Simultaneous use of weak and strong variations.** Suppose again that we are given q^* admissible weak variations $[\zeta_q(t, r), \eta_q(1)]$ ($q = 1, \cdots, q^*$), and that we are also given several admissible strong variations. We denote them by $[t_q, r_q]$ ($q = q^* + 1, \cdots, \bar{q}$) and suppose the numbering so chosen that

$$0 \leqq t_{q^*+1} \leqq t_{q^*+2} \leqq \cdots \leqq t_{\bar{q}} \leqq 1. \tag{7.1}$$

We apply the above construction $\bar{q} - q^*$ times. First, using the strong variations $[t_q, r_q]$, $q = q^* + 1$, we construct the curve analogous to $y = Y(s, b)$ in §6; here we denote it by

$$y = Y_{q^*+1}(t, b). \tag{7.2}$$

Next we construct the analogue of $\bar{C}^*_{b,\tau}$, with τ replaced by the symbol b_{q^*+1}; equation (6.5) becomes

$$\bar{y}^i(t_{q^*+1}, b) = Y_{q^*+1}(b), \qquad b = (b, \cdots, b_{q^*+1}). \tag{7.3}$$

The curve C^*_b, $b = (b_1, \cdots, b_{p^*+1})$, is now defined as in (6.6). Next we use the second of the collection of strong variations, $[t_q, r_q]$ with $q = q^* + 2$. In constructing the analogue of $Y(s, b)$, or (7.2), we use the tracks $\bar{y}(t, b)$ of the auxiliary curves $\bar{C}^*_b$ ($b = (b_1, \cdots, b_{q^*+1})$), which at $t = t_{q^*+2}$ are also the tracks of the C^*_b. We obtain a new set of auxiliary curves $\bar{C}^*_b$ and a new family of generalized curves C^*_b, where b now means $(b, \cdots, b_{q^*+2})$. This curve C^*_b consists of five arcs: first an arc of $\bar{C}^*_b$ with q^* parameters b, then an arc of ordinary curve (7.2) with q^* parameters b, then an arc $\bar{C}^*_b$ with $q^* + 1$ parameters b, then an arc of ordinary curve $y = Y_{q^*+2}(t, b, \cdots, b_{q^*+1})$, and finally an arc of $\bar{C}^*_b$ with $q^* + 2$ parameters b.

Repeating the process, we obtain a family of generalized curves $\bar{C}^*_b$, $b = (b_1, \cdots, b_{\bar{q}})$. The derivatives

$$\frac{\partial}{\partial b_q} \mathcal{F}(C^*_b), \qquad \frac{\partial}{\partial b_q} y(0, b), \qquad \frac{\partial}{\partial b_q} y(1, b)$$

are all defined and continuous for all b near 0, if we understand $\mathcal{F}(C^*_b)$ to be defined formally by equation (6.7). It must, however, be remembered that for this to coincide with the ordinary meaning of $\mathcal{F}(C^*_b)$ we must suppose, as indeed we have supposed, that $b_q \geqq 0$ for $q = q^* + 1, \cdots, \bar{q}$; for otherwise in the second term on the right in (6.7) we are integrating backwards along an arc. That is, for $\tau < 0$ the second integral in (6.7) represents the negative of the integral along the reversed arc of $y = Y(s, b)$ from $s = 0$ to $s = \tau$ instead of the integral along the arc from $s = \tau$ to $s = 0$. The value of the partial

derivative of $J(C_b^*)$ for $b = 0$ is computed from (5.10) if $1 \leqq q \leqq q^*$, while if $q^* < q \leqq \bar{q}$ we have by (6.9)

$$(7.4) \quad \lambda^0 \frac{\partial}{\partial b_q} J(C_b^*)\Big|_{b=0} = G(0, \eta_q(1)) + h^i(1)\eta_q^i(1) + \{\lambda^0 f(y_0(t_q), r_q) - r_q^i h^i(t_q)\}.$$

8. **A convex set defined by the variations.** Each admissible weak variation $[\zeta(t, r), \eta(1)]$ determines an embedding family, and thereby determines a first variation

$$\frac{d}{db} J(C_b^*)\Big|_{b=0}$$

of the integral. It also determines the derivatives of each end function ψ^μ, for

$$(8.1) \quad \frac{d}{db}\psi^\mu(y(0, b), y(1, b))\Big|_{b=0} = \Psi^\mu(\eta(0), \eta(1)) \qquad (\mu = 1, \cdots, p),$$

where we define

$$(8.2) \quad \Psi^\mu(\eta_1, \eta_2) = \psi^\mu_{y_1^i}(y_0(0), y_0(1))\eta_1^i + \psi^\mu_{y_2^i}(y_0(0), y_0(1))\eta_2^i.$$

Thus each admissible weak variation $[\zeta(t, r), \eta(1)]$ determines a vector in $(p + 1)$-dimensional space with components

$$(8.3) \quad \left(\frac{d}{db} J(C_b^*)\Big|_{b=0}, \Psi^1(\eta(0), \eta(1)), \cdots, \Psi^p(\eta(0), \eta(1))\right).$$

Likewise, each admissible strong variation determines a family of curves as described in §§6, 7, and thereby determines a vector

$$(8.4) \quad \left(\frac{d}{db} J(C_b^*)\Big|_{b=0}, \Psi^1(0, \eta(1)), \cdots, \Psi^p(0, \eta(1))\right).$$

Consider the aggregate V of all points $(u^0, \cdots, u^p)$ in a $(p + 1)$-dimensional space of points u which are given by (8.3) or (8.4). Let δ_0 be the vector $(1, 0, \cdots, 0)$. The aggregate K of all linear combinations $b_0\delta_0 + b_1v_1 + \cdots + b_sv_s$ of δ_0 and vectors v_i of V with positive coefficients b_i is easily seen to be a convex point set. So is its closure $\bar{K}$. We now prove

LEMMA 8.1. *The origin of u-space is not interior to* $\bar{K}$.

Suppose the origin $u = (0, \cdots, 0)$ is interior to $\bar{K}$. If we choose a simplex with vertices $u_1, \cdots, u_{p+2}$ sufficiently near $(0, \cdots, 0)$, these vertices will belong to $\bar{K}$. We suppose the simplex so chosen that it contains the origin in its interior.

Arbitrarily near each vertex of the simplex there are points of K. We may therefore choose $p + 2$ points of K which are near enough to the vertices of the original simplex so that they themselves are the vertices of a simplex containing the origin in its interior. There is no loss of generality in supposing that the u_k $(k = 1, \cdots, p + 2)$ were already so chosen.

Since the origin is interior to the simplex, there are positive numbers $w_1, \cdots, w_{p+2}$ such that $\sum w_k = 1$ and

$$w_k u_k^i = 0 \qquad (i = 0, \cdots, p;\ k = 1, \cdots, p + 2). \tag{8.5}$$

By definition of K, each vector u_k is a sum of the form (suspending the summation convention)

$$u_k^i = \bar{b}_{k,0}\delta_0^i + \sum_{h=s_{k-1}}^{s_k-1} \bar{b}_h v_h^i, \tag{8.6}$$

in which the v_h belong to V, the coefficients $\bar{b}_h$ and $\bar{b}_{k,0}$ are all positive, and $0 = s_0 \leqq s_1 \leqq \cdots \leqq s_{p+2}$. If we substitute these in (8.5) and define

$$\bar{b}_h = w_k\bar{b}_h / \sum_1^{p+2} \bar{b}_{k,0} \qquad (h = s_{k-1}, \cdots, s_k - 1;\ k = 1, \cdots, p + 2),$$

we obtain (restoring the summation convention)

$$-\delta_0^i = \bar{b}_h v_h^i \qquad (h = 0, 1, \cdots, s = s_{p+2} - 1;\ i = 0, \cdots, p). \tag{8.7}$$

The rows of the matrix $\| v_h^i \|$ in (8.7) cannot be linearly dependent. If they were, there would exist numbers $l_0, \cdots, l_p$ such that

$$l_i v_h^i = 0 \qquad (h = 0, \cdots, s).$$

By (8.7) and (8.6) this would give

$$w_k u_k^i l_i = 0$$

identically in $w_1, \cdots, w_{p+2}$. But this is impossible; no such linear relation can hold, because as the w_k vary over all positive numbers the points $w_k u_k$ include the whole interior of the simplex with vertices u_k. Thus we have shown that at least one $(p + 1)$-square minor of $\| v_h^i \|$ is non-singular. By renumbering if necessary, we can bring it about that the minor

$$\| v_h^i \| \qquad (i, h = 0, 1, \cdots, p) \tag{8.8}$$

is non-singular.

Each of the vectors v_h $(h = 0, \cdots, s)$ arises either from an admissible weak variation $[\zeta(t, r), \eta(1)]$ or from an admissible strong variation $[t_0, r_0]$. We construct as in §7 a family of curves C_b^*, where b is $(b_0, \cdots, b_s)$, such that at $b = 0$ the vector

$$\left(\frac{\partial}{\partial b_h} J(C_b^*), \frac{\partial}{\partial b_h}\psi^1, \cdots, \frac{\partial}{\partial b_h}\psi^p\right)$$

is v_h. The equations

$$\begin{aligned} &J(C_b^*) - J(C_0^*) + u = 0, \\ &\psi^\mu(y(0, b), y(1, b)) = 0 \qquad (\mu = 1, \cdots, p) \end{aligned} \tag{8.9}$$

have the initial solution $u = b = 0$, as remarked in §7. The Jacobian with respect to $b_0, \cdots, b_p$ is non-vanishing for $u = b = 0$, being the determinant of the matrix (8.8). Hence the equations have solutions

$$b_h = b_h(u, b_{p+1}, \cdots, b_s) \qquad (h = 0, 1, \cdots, p),$$

which are continuous and continuously differentiable near $u = b_{p+1} = \cdots = b_s = 0$ and there reduce to $b_h = 0$. With the constants $\bar{b}_k$ of (8.7) we now define

$$\begin{aligned} b_k(\tau) &= \bar{b}_k\tau && (k = p+1, \cdots, s), \\ b_h(\tau) &= b_h(\tau, b_{p+1}(\tau), \cdots, b_s(\tau)) && (h = 0, 1, \cdots, p). \end{aligned} \tag{8.10}$$

Then (8.9) yields the identities

$$\begin{aligned} &J(C^*_{b(\tau)}) - J(C^*_0) + \tau = 0, \\ &\psi^\mu(y(0, b(\tau)), y(1, b(\tau))) = 0 \qquad (\mu = 1, \cdots, p). \end{aligned} \tag{8.11}$$

If we differentiate with respect to τ and set $\tau = 0$, we obtain, in v_h-notation,

$$\begin{aligned} &v_0^0 b_0'(0) + \cdots + v_p^0 b_p'(0) + v_{p+1}^0 \bar{b}_{p+1} + \cdots + v_s^0 \bar{b}_s = -1, \\ &v_0^i b_0'(0) + \cdots + v_p^i b_p'(0) + v_{p+1}^i \bar{b}_{p+1} + \cdots + v_s^i \bar{b}_s = 0 \quad (i = 1, \cdots, p). \end{aligned} \tag{8.12}$$

The matrix of coefficients of the $b_h'(0)$ is non-singular, so on comparing (8.12) with (8.7) we obtain

$$b_h'(0) = \bar{b}_h > 0 \qquad (h = 0, 1, \cdots, p). \tag{8.13}$$

By choosing τ to be an arbitrarily small positive number, the track of $C^*_{b(\tau)}$ can be made to have an arbitrarily small Fréchet distance from the track of C^*_0. Moreover, for such τ all the numbers $b_h(\tau)$ will be positive, the first $p + 1$ of them because of (8.13) and the rest because of the definition (8.10). Hence $\mathcal{F}(C^*_{b(\tau)})$ will be the integral of f along $C^*_{b(\tau)}$ in the usual sense. By (8.11), the curves $C^*_{b(\tau)}$ satisfy the end conditions $\psi^\mu = 0$ and give to $J(C^*)$ a value smaller than $J(C^*_0)$. This contradicts hypothesis (1.7), and our lemma is established.

9. **Derivation of the multiplier rule.** By Lemma 8.1, the origin in $(u^0, \cdots, u^p)$-space is not interior to $\bar{K}$. However, it is obvious that the origin belongs to $\bar{K}$. Hence it is a boundary point of $\bar{K}$, and through it there passes a hyperplane of support of $\bar{K}$. That is, there is a linear function

$$L(u) \equiv \lambda^0 u^0 + e_\mu u^\mu \qquad (\mu = 1, \cdots, p;\ \lambda^0, e_\mu \text{ not all zero}) \tag{9.1}$$

such that $L(u) \geqq 0$ for every point u of $\bar{K}$. In particular, δ_0 belongs to $\bar{K}$, so that

$$\lambda^0 = L(\delta_0) \geqq 0. \tag{9.2}$$

We suppose that the λ^0 of Lemma 2.1 is chosen to be the λ^0 of (9.1); the c_i will be chosen shortly. Consider now an admissible strong variation $[t_0, r_0]$.

This gives a vector $(u^0, \cdots, u^p)$ defined by (6.9) and (8.4), and for this vector the inequality $L(u) \geqq 0$ takes the form

$$\lambda^0 G(0, \eta(1)) + h^i(1)\eta^i(1) + e_\mu \psi^\mu(0, \eta(1)) + \{\lambda^0 f(y_0(t_0), r_0) - r_0^i h^i(t_0)\} \geqq 0. \tag{9.3}$$

We shall simplify this later. Next, an admissible weak variation $[\zeta(t, r), \eta(1)]$ gives a vector $(u^0, \cdots, u^p)$ defined by (5.10) and (8.3). For this vector the inequality $L(u) \geqq 0$ becomes

$$\begin{aligned}&\lambda^0 G(\eta(0), \eta(1)) + e_\mu \Psi^\mu(\eta(0), \eta(1)) + \eta^i(t)h^i(t)\Big|_0^1 \\ &\qquad - \int_0^1 \mathfrak{M}_0[t; \lambda^j(t, r)\zeta^j(t, r)]\, dt \geqq 0.\end{aligned} \tag{9.4}$$

However, $[-\zeta(t, r), -\eta(1)]$ is also an admissible variation; and for this the left member of (9.4) changes sign, the inequality remaining valid. This is only possible if equality holds in (9.4).

The number $\eta^i(0)$ occurs several times in (9.4), and in one occurrence it has the coefficient $-h^i(0) = -c_i$. Hence it is possible to choose the c_i in Lemma 2.1 in such a way that the coefficient of $\eta^i(0)$ in (9.4) (which is the left member of the first of equations (9.5) below) is zero. We suppose the c_i so chosen. Next, we set $\zeta^j(t, r) \equiv 0$ $(j = 1, \cdots, \nu)$. The left member of (9.4) reduces to a linear function of the $\eta^i(1)$. Since the $\eta^i(1)$ can be chosen arbitrarily, the identical vanishing of the linear function implies that all its coefficients are zero. Thus far we have shown

$$\begin{aligned}&\lambda^0 g_{y_1^i}(y_0(0), y_0(1)) + e_\mu \psi^\mu_{y_1^i}(y_0(0), y_0(1)) - c_i = 0, \\ &\lambda^0 g_{y_2^i}(y_0(0), y_0(1)) + e_\mu \psi^\mu_{y_2^i}(y_0(0), y_0(1)) + h^i(1) = 0 \qquad (i = 1, \cdots, \nu).\end{aligned} \tag{9.5}$$

By virtue of (9.5), equation (9.4) reduces to

$$\int_0^1 \mathfrak{M}_0[t; \lambda^j(t, r)\zeta^j(t, r)]\, dt = 0. \tag{9.6}$$

If we choose

$$\begin{aligned}\zeta^\alpha(t, r) &= 0 \qquad &(\alpha = i, \cdots, m),\\ \zeta^\beta(t, r) &= \lambda^\beta(t, r) \qquad &(\beta = m + 1, \cdots, \nu),\end{aligned}$$

equations (9.6) take the form

$$\int_0^1 \mathfrak{M}_0[t; \lambda^\beta(t, r)\lambda^\beta(t, r)]\, dt = 0. \tag{9.7}$$

It follows at once that for almost all t in M_0 we have

$$\mathfrak{M}_0[t; \lambda^\beta(t, r)\lambda^\beta(t, r)] = 0;$$

whence, by definition, no vector r_0 such that any $\lambda^\beta(t, r)$ is non-vanishing can be carried at $y_0(t_0)$. Therefore

(9.8) *for almost all t in M_0, all the functions $\lambda^\beta(t, r)$ $(\beta = m + 1, \cdots, \nu)$ vanish at each vector r_0 carried by C_0^* at $y_0(t_0)$.*
By rejecting a set of measure zero from M_0, we can bring it about that (9.8) holds for all t in M_0.

If now we replace the functions $\lambda^\beta(t, r)$ $(\beta = m + 1, \cdots, \nu)$ by 0, the left member of (2.2) remains unchanged at each vector carried. So does the integral in the right member, by Lemma 10.1 of GC; and therefore
(9.9) *if we set $\lambda^\beta = 0$, equation (2.2) still holds for almost all t in M_0, provided that the vectors r in the left member are carried at $y_0(t_0)$.*

The functions $\lambda^\alpha(t, r)$, as pointed out in Lemma 2.1, are continuous in (t, r) and of class $C^{\kappa-1}$ in r for each fixed t, provided that $(y_0(t), r)$ is in $\bar{U}$. It is in some respects desirable, though not vital, to extend the range of definition of $\lambda^\alpha(t, r)$ to all (t, r), preserving the continuity properties described. The possibility of this extension will be proved in §18 of this paper.

10. **The principal theorem.** In defining the Weierstrass E-function it is convenient to depart slightly from the usual notation; the departure will not affect the meaning. Let $r, \bar{r}$ be vectors such that (y, r) and $(y, \bar{r})$ are in R, and $|r| \neq 0$. For arbitrary numbers $\lambda^0, \lambda^1, \cdots, \lambda^m$ we define

$$(10.1)\quad \begin{aligned} E(y, r, \lambda, \bar{r}) &= \lambda^0 f(y, \bar{r}) - \bar{r}^i\{\lambda^0 f_{r^i}(y, r) + \lambda^\alpha \varphi^\alpha_{r^i}(y, r)\} \\ &= \lambda^0 f(y, \bar{r}) - \bar{r}^i F_{r^i}(y, r, \lambda). \end{aligned}$$

Our principal theorem is

THEOREM 10.1. *Let the hypotheses of §1 be satisfied. Then there exist a non-negative constant λ^0, a set of functions $\lambda^\alpha(t, r)$ $(\alpha = 1, \cdots, m)$ defined and continuous for all t and r, and a subset M of the interval* [0, 1] *with measure $mM =$ 1, such that if $F(y, r, \lambda)$ is defined by* (2.1) *the following conclusions hold.*

(I) (*DuBois-Reymond relations.*) *There are constants $c_1, \cdots, c_\nu$ with which the equations*

$$(10.2)\quad F_{r^i}(y_0(t), r, \lambda(t, r)) = c_i + \int^t \mathfrak{M}_0[t; F_{y^i}(y_0(t), r, \lambda(t, r))]\, dt$$

are satisfied for all t in M and all vectors r carried by C_0^* at $y_0(t)$.

(Ia) (*Transversality condition.*) *There are numbers $e_1, \cdots, e_\mu$ such that the equations*

$$(10.3)\quad \lambda^0 g_{y^i_s}(y_0(0), y_0(1)) + e_\mu \psi^\mu_{y^i_s}(y_0(0), y_0(1)) + (-1)^{s+1} h^i(s) = 0 \qquad (s = 0, 1;\ i = 1, \cdots, \nu)$$

are satisfied, where $h^i(t)$ is merely an abbreviation for the right member of (10.2).

(II) (*Weierstrass condition, first form.*) *For all t_0 in* [0, 1] *and all vectors r such that $(y_0(t_0), r)$ is admissible, the inequality*

$$(10.4)\quad \lambda^0 f(y_0(t_0), r) - h^i(t_0) r^i \geqq 0$$

is satisfied.

(*Weierstrass condition, second form.*) *For all* t_0 *in* M, *all vectors* r_0 *carried by* C_0^* *at* $y_0(t_0)$, *and all vectors* r *such that* $(y_0(t_0), r)$ *is admissible, the inequality*

$$E(y_0(t_0), r_0, \lambda(t_0, r_0), r) \geqq 0 \tag{10.5}$$

is satisfied.

(III) (*Clebsch condition.*) *If* $f(y, r)$ *and the* $\varphi^\alpha(y, r)$ *are of class* C^2, *then whenever* t_0 *is in* M *and* r_0 *is carried at* $y_0(t_0)$ *the inequality*

$$F_{r^i r^j}(y_0(t_0), r_0, \lambda(t, r_0))\pi^i\pi^j \geqq 0 \tag{10.6}$$

holds for every set of numbers π^i $(i = 1, \cdots, \nu)$ *such that*

$$\varphi^\alpha_{r^i}(y_0(t_0), r_0)\,\pi^i = 0 \qquad (\alpha = 1, \cdots, m). \tag{10.7}$$

Moreover, for all t_0 *in* M *and all* r_0 *carried by* C_0^* *at* $y_0(t_0)$ *the inequality*

$$|\lambda^0| + \sum_{\alpha=1}^{m} |\lambda^\alpha(t_0, r_0)| > 0 \tag{10.8}$$

is satisfied. The numbers λ^0, e_μ *are not all zero. For all fixed* t, *the functions* $\lambda^\alpha(t, r)$ *are of class* $C^{\kappa-1}$ *in* r (cf. hypothesis (1.2)).

Conclusion (I) is merely a restatement of (9.9). Conclusion (Ia) is a restatement of equations (9.5), since $c_i = h^i(0)$. In inequality (9.3), the terms involving $\eta(1)$ vanish by conclusion (Ia); thus (9.3) takes the form (10.4). This, with (10.1), conclusion (I) and the definition of $h(t)$, yields (10.5).

As usual, we deduce the condition of Clebsch from that of Weierstrass. The numbers t_0, r_0 and π being as described in (III), we write the equations

$$\begin{aligned} \varphi^\alpha(y_0(t_0), r) &= 0 & (\alpha &= 1, \cdots, m), \\ \varphi^\beta(y_0(t_0), r) - \tau\varphi^\beta_{r^i}(y_0(t_0), r_0)\pi^i &= 0 & (\beta &= m + 1, \cdots, n). \end{aligned} \tag{10.9}$$

Since the Jacobian of the left members with respect to the r is the non-vanishing determinant in (1.9), these have solutions $r^i = r^i(\tau)$ reducing to r_0 for $\tau = 0$ and of class C^2 for τ near 0. If we differentiate in (10.9), we obtain

$$\varphi^\alpha_{r^i}(y_0(t_0), r_0)r^{i\prime}(0) = 0 = \varphi^\alpha_{r^i}(y_0(t_0), r_0)\pi^i,$$
$$\varphi^\beta_{r^i}(y_0(t_0), r_0)r^{i\prime}(0) = \varphi^\beta_{r^i}(y_0(t_0), r_0)\pi^i.$$

The matrix of coefficients being non-singular, we must have

$$r^{i\prime}(0) = \pi^i. \tag{10.10}$$

By (10.9), the element $(y_0(t_0), r(\tau))$ is admissible, so by (II) we have

$$\gamma(\tau) \equiv \lambda^0 f(y_0(t_0), r(\tau)) - r^i(\tau)F_{r^i}(y_0(t_0), r_0, \lambda(t_0, r_0)) \geqq 0$$

for all τ near zero. Because of homogeneity, $\gamma(\tau)$ vanishes at $\tau = 0$, so it has a minimum there. Hence

$$\begin{aligned} \gamma''(0) = \lambda^0 f_{r^i r^j}(y_0(t_0), r_0)\pi^i\pi^j &+ \lambda^0 f_{r^i}(y_0(t_0), r_0)r^{i\prime\prime}(0) \\ &- r^{i\prime\prime}(0)F_{r^i}(y_0(t_0), r_0, \lambda(t_0, r_0)) \geqq 0. \end{aligned} \tag{10.11}$$

A second differentiation in (10.9) yields

$$\varphi^{\alpha}_{r^i r^j}(y_0(t_0), r_0)\pi^i\pi^j + \varphi^{\alpha}_{r^i}(y_0(t_0), r_0)r^{i''}(0) = 0. \tag{10.12}$$

If we multiply by $\lambda^{\alpha}(t_0, r_0)$ and add member by member to (10.11), we obtain[14] (10.6) for τ near zero.

It remains to establish inequality (10.8). Suppose that this is false, and that for some t_0 in M and some r_0 carried at $y_0(t_0)$ all $m + 1$ terms on the left of (10.8) vanish. We then have

$$F_{r^i}(y_0(t_0), r_0, \lambda(t_0, r_0)) = 0.$$

Since t_0 is in M, equations (10.2) hold at t_0, so $h^i(t_0) = 0$ $(i = 1, \cdots, \nu)$. The constant λ^0 being zero, equations (2.5) are homogeneous linear differential equations for the $h^i(t)$; and since the $h^i(t)$ satisfy these equations and vanish at t_0, they vanish identically. Equations (10.3) now express a linear dependence among the rows of the matrix

$$\| \psi^{\mu}_{y_1^i}(y_0(0), y_0(1)), \qquad \psi^{\mu}_{y_2^i}(y_0(0), y_0(1)) \|,$$

contradicting hypothesis (1.10). Hence conclusion (10.8) is established.

11. **Geometric interpretation of the Weierstrass condition.** The Weierstrass condition can be given a geometric interpretation. For each fixed t we consider the surface in the $(\nu + 1)$-dimensional space of variables $(r^1, \cdots, r^{\nu}, u)$ defined by the equation

$$u = \lambda^0 f(y_0(t), r). \tag{11.1}$$

Also, we define a hyperplane by the equation

$$u = L(t, r) = h^i(t)r^i. \tag{11.2}$$

Then from inequality (10.4), equations (10.2) and the homogeneity of F we find at once the following corollary.

COROLLARY 11.1. *If t_0 is in $[0, 1]$, the surface (11.1) does not lie below the hyperplane (11.2) for any r such that $(y_0(t_0), r)$ is admissible, while if t_0 is in M and r_0 is carried by C_0^* at $y_0(t_0)$ the hyperplane and the surface meet at $r = r_0$.*

12. **Change of parameter.** We have not imposed any homogeneity conditions on the $\lambda^{\alpha}(t, r)$. However, since C_0^* is represented in terms of standard parameter, all vectors r carried by C_0^* have length $| r | = \mathfrak{L}(C_0^*)$. Thus if we define for $| r | > 0$

$$\bar{\lambda}^{\alpha}(t, r) = \lambda^{\alpha}(t, r\mathfrak{L}(C_0^*)/| r |),$$

the functions $\bar{\lambda}^{\alpha}$ are continuous for all t and for all $r \neq 0$, and are positively homogeneous of degree 0 in r for fixed t. Moreover, $\bar{\lambda}^{\alpha}$ and λ^{α} are equal when-

[14] This proof was suggested by Professor Graves, and replaces one which used the differentiability of $\lambda^{\alpha}(t, r)$ as a function of r.

ever t is in M and r is carried at $y_0(t_0)$, so that we may replace the λ^α by the $\bar{\lambda}^\alpha$ in Theorem 10.1 without injuring any of the conclusions. Thus, if we are willing to permit the $\lambda^\alpha(t, r)$ to be discontinuous at $r = 0$ (and this is harmless), we may suppose that they are positively homogeneous of degree 0 in r.

Under this assumption, $F(y, r, t, \lambda)$ is positively homogeneous of degree 1 in r, and we have the privilege of changing parameter from t to any parameter $\tau = \tau(t)$ such that $\tau'(t)$ is absolutely continuous and $\tau'(t)$ is positive for almost all t, as in §4 of GC.

13. **Supporting sets.** In the next section, and again in the third paper of this sequence, we shall need the concept of a supporting set. This is the extension to Bolza problems of the concept of "$\mathfrak{S}$-admissible approach set" defined in an earlier paper,[15] but some slight difficulties are introduced by the dependence of the multipliers λ^α on the variables r.

Let y_0 be a fixed point, and let $\lambda^0 \geqq 0, l_1, \cdots, l_\nu$ be fixed numbers. If the expression

$$e(r) \equiv \lambda^0 f(y_0, r) - r^i l_i \tag{13.1}$$

is negative for some r such that (y_0, r) is admissible, we say that (λ^0, l) determines no supporting set. If (13.1) is non-negative for such r, we define the supporting set $S[y_0, \lambda^0, l]$ to be the set of all vectors r such that (y_0, r) is admissible and the equation

$$\lambda^0 f(y_0, r) - r^i l_i = 0 \tag{13.2}$$

holds. It follows that if (y_0, r) is admissible and r is not in $S[y_0, \lambda^0, l]$, the inequality

$$\lambda^0 f(y_0, r) - r^i l_i > 0 \tag{13.3}$$

must hold.

It is clear that if r_1 is a non-null vector in $S[y_0, \lambda^0, l]$, the function $e(r)$ has its minimum value 0 at $r = r_1$ subject to the conditions $\varphi^\alpha(y_0, r) = 0$. Hence by the multiplier rule for functions of a finite number of variables, there are numbers $\Lambda, \lambda^\alpha(r_1)$ not all zero such that all the first partials of the combination

$$\Lambda e(r) + \lambda^\alpha(r_1)\varphi^\alpha(y_0, r) \tag{13.4}$$

vanish at r_1. If $\Lambda = 0$, this contradicts the assumption that the matrix in (1.9) has rank m at (y_0, r_1); hence Λ is not zero, and we may choose it to be 1. The other multipliers $\lambda^\alpha(r_1)$ are then uniquely determined. The vanishing of the partial derivatives of the combination (13.4) yields the equations

$$\lambda^0 f_{r^i}(y_0, r_1) + \lambda^\alpha(r_1)\varphi^\alpha_{r^i}(y_0, r_1) = l_i . \tag{13.5}$$

[15] E. J. McShane, *The isoperimetric problem in parametric form*, Trans. Amer. Math. Soc., vol. 45(1939), pp. 197–216; in particular, p. 212.

The functions $\lambda^\alpha(r)$ thus uniquely determined on $S[y_0, \lambda^0, l]$ will be called the *multipliers for* $S[y_0, \lambda^0, l]$. From this definition and equation (13.5) we deduce

LEMMA 13.1. *Let $S[y_0, \lambda^0, l]$ be a supporting set, and let $\lambda^1(r), \cdots, \lambda^m(r)$ be its multipliers. If as usual we define $F(y, r, \lambda) = \lambda^0 f(y, r) + \lambda^\alpha(r)\varphi^\alpha(t, r)$, then the equations*

$$F_{r^i}(y_0, r, \lambda) = l_i$$

hold for all r in $S[y_0, \lambda^0, l]$.

From Corollary 11.1 we readily deduce the following lemma.

LEMMA 13.2. *Let conclusions* (I) *and* (II) *of Theorem* 10.1 *hold with multipliers λ^0, $\lambda^\alpha(t, r)$. Then for each t in M there is a supporting set $S(t) \equiv S[y_0(t), \lambda^0, h(t)]$ which contains all vectors r carried by C_0^* at $y_0(t)$. The multipliers $\lambda^\alpha(r)$ of the set $S(t)$ are identical with the multipliers $\lambda^\alpha(t, r)$ for all vectors r carried by C_0^* at $y_0(t)$.*

Corollary 11.1 states that the supporting set $S(t)$ exists and that each vector r carried by C_0^* at $y(t)$ is contained in S. (The proof of this corollary used only conclusions (I) and (II) of Theorem 10.1.) Equations (10.2) inform us that $\lambda^\alpha(t, r)$ satisfies equation (13.5). The solutions of the latter being unique, we must have $\lambda^\alpha(t, r_1) = \lambda^\alpha(r_1)$ for every vector r carried at $y_0(t)$. This completes the proof of the lemma.

Of course it is possible that there may be several independent sets of multipliers λ^0, $\lambda^\alpha(t, r)$ with which Theorem 10.1 holds. We are supposing that one such set has been chosen; the supporting sets $S(t)$ are then uniquely determined for all t in M. It is conceivable that a different choice of multipliers (if possible) might have led to a different supporting set $S(t)$.

14. **Analogue of the Dresden corner condition.** For an arbitrary constant λ^0 and arbitrary functions $\lambda^\alpha(r)$ we define the Carathéodory Ω-function by the equation

$$\Omega(y, r, \bar{r}, \lambda) = r^i F_{y^i}(y, \bar{r}, \lambda(\bar{r})) - \bar{r}^i F_{y^i}(y, r, \lambda(r)). \tag{14.1}$$

We now proceed to establish a condition involving the Ω-function which may be regarded as a rather far-fetched analogue of Dresden's corner condition. In this theorem we shall understand the $\lambda^\alpha(t, r)$ to be the multipliers for the supporting set $S(t)$ of Lemma 13.2; these coincide with the multipliers of Theorem 10.1 if r is carried at $y_0(t)$, and are defined on all of $S(t)$, even for r not carried.

THEOREM 14.1. *Let conclusions* (I) *and* (II) *of Theorem* 10.1 *be satisfied with multipliers λ^0, $\lambda^\alpha(t, r)$. Then for almost all t_0 in M_0 the equation*

$$\mathfrak{M}_0[t_0\,; \Omega(y_0(t_0), r_0, r, \lambda)] = 0 \tag{14.2}$$

holds for all vectors r_0 in the supporting set $S(t_0)$.

By rejecting a set of measure zero from the set M of Theorem 10.1, we may bring it about that for all t in M the functions $y_0^i(t)$ and $h^i(t)$ are differentiable and the derivative of h^i is the integrand in (10.2). Let t_0 be a point in M, and r_0 a vector of $S(t_0)$.

Since r_0 is in $S(t_0)$, the set $(y_0(t_0), r_0)$ is admissible. It is therefore possible to adjoin functions $\varphi^\beta(y, r)$ $(\beta = m + 1, \cdots, \nu)$, defined and of class C′ near $(y_0(t_0), r_0)$, such that the matrix

$$\| \varphi^i_{r^j}(y, r) \| \qquad (i, j = 1, \cdots, \nu)$$

is non-singular at $(y_0(t_0), r_0)$. By the implicit functions theorem, the equations

$$\varphi^\alpha(y, r) = 0, \qquad \varphi^\beta(y, r) = \varphi^\beta(y_0(t_0), r_0) \tag{14.3}$$

have solutions $r^i = \bar{r}^i(y)$ defined and of class C′ near $y_0(t_0)$, and reducing to r_0 for $y = y_0(t_0)$. The equations

$$\varphi^\alpha(y_0(t), \bar{r}(y_0(t))) = 0 \qquad (\alpha = 1, \cdots, m)$$

are identities for t near t_0, so by differentiating and setting $t = t_0$ we find

$$\varphi^\alpha_{y^i}(y_0(t_0), r_0)\dot{y}_0^i(t_0) + \varphi^\alpha_{r^i}(y_0(t_0), r_0)\bar{r}^i_{y^j}\dot{y}_0^j(t_0) = 0. \tag{14.4}$$

By conclusion (II) of Theorem 10.1, the inequality

$$\lambda^0 f(y_0(t), \bar{r}(y_0(t))) - \bar{r}^i(y_0(t))h^i(t) \geqq 0 \tag{14.5}$$

holds for all t near t_0. For $t = t_0$, the vector $r_0 = \bar{r}(y_0(t_0))$ is in $S(t_0)$, hence the left member of (14.5) vanishes. Thus the left member of (14.5) has a minimum at t_0, and its derivative must vanish at t_0:

$$\begin{aligned} \lambda^0 f_{y^i}(y_0(t_0), r_0)\dot{y}_0^i(t_0) &+ \lambda^0 f_{r^i}(y_0(t_0), r_0)\bar{r}^i_{y^j}(y_0(t_0))\dot{y}_0^j(t_0) \\ &- \bar{r}^i_{y^j}(y_0(t_0))\dot{y}_0^j(t_0)h^i(t_0) - r_0^i\dot{h}^i(t_0) = 0. \end{aligned} \tag{14.6}$$

By conclusion (I) of Theorem 10.1 and the choice of t_0, we have

$$h^i(t_0) = \lambda^0 f_{r^i}(y_0(t_0), r_0) + \lambda^\alpha(r_0)\varphi^\alpha_{r^i}(y_0(t_0), r_0),$$
$$\dot{h}^i(t_0) = \mathfrak{M}_0[t_0\,; F_{y^i}(y_0(t_0), r, \lambda(r))].$$

We substitute these in (14.6); recalling (14.4), we find

$$\dot{y}_0^i(t_0)F_{y^i}(y_0(t_0), r_0, \lambda(r_0)) - r_0^i\mathfrak{M}_0[t_0\,; F_{y^i}(y_0(t_0), r, \lambda(r))] = 0. \tag{14.7}$$

This, with (2.8c) of GC, implies

$$\mathfrak{M}_0[t_0\,; r^iF_{y^i}(y_0(t_0), r_0, \lambda(r_0)) - r_0^iF_{y^i}(y_0(t_0), r, \lambda(r))] = 0. \tag{14.8}$$

This is conclusion (14.2) of our theorem.

15. Isoperimetric conditions. Let us suppose that to the side conditions $\varphi^\alpha(y, r) = 0$ we add several side conditions of isoperimetric type:

$$\mathcal{G}^\gamma(C^*) = \int_a^b \mathfrak{M}[t; g^\gamma(y(t), r)]\, dt = l^\gamma \qquad (\gamma = 1, \cdots, q), \tag{15.1}$$

where the g^γ satisfy the usual continuity and homogeneity requirements (1.2) and the l^γ are constants. By the usual device we reduce conditions (15.1) to Bolza form. We introduce new variables $y^{\nu+\gamma}$, $r^{\nu+\gamma}$ ($\gamma = 1, \cdots, q$), and impose the differential equations

$$r^{\nu+\gamma} + g^\gamma(y, r) = 0 \tag{15.2}$$

and the end conditions

$$y^{\nu+\gamma}(a) = 0, \qquad y^{\nu+\gamma}(b) = -l^\gamma. \tag{15.3}$$

Conditions (15.2) and (15.3) are equivalent to (15.1).

The new rows introduced in the matrices in (1.8) and (1.9) are independent of the old ones, so admissibility requirements and hypothesis (1.8) are unaltered. The definition (2.1) of F now becomes

$$\begin{aligned} F(y, r, \lambda(t, r)) = \lambda^0 f(y, r) + \lambda^\alpha(t, r)\varphi^\alpha(t, r) \\ + \lambda^{\nu+\gamma}(t, r)\{r^{\nu+\gamma} + g^\gamma(y, r)\}. \end{aligned} \tag{15.4}$$

Equations (10.2) are augmented by q new ones,

$$\lambda^{\nu+\gamma}(t, r) = c_{\nu+\gamma} + \int_0^t 0\, dt = c_{\nu+\gamma}. \tag{15.5}$$

These hold for all t in M and all r carried at $y_0(t)$. Hence we may replace $\lambda^{\nu+\gamma}(t, r)$ by the constant $c_{\nu+\gamma}$ without altering any conclusions in Theorem 10.1 or later theorems; that is, we may write $\lambda^{\nu+\gamma}$ (a constant) in place of $\lambda^{\nu+\gamma}(t, r)$.

In (15.4) the terms $\lambda^{\nu+\gamma}r^{\nu+\gamma}$ are readily seen to be insignificant. Their omission leaves the E-function and the quadratic form in (10.6) unaltered, and likewise makes no change in the first ν of equations (10.2) and (10.3). Hence we may omit them from (15.4), obtaining the familiar result that to isoperimetric conditions (15.1) there correspond terms $\lambda^\gamma g^\gamma(y, r)$ in the sum F, the conclusions of Theorem 10.1 and subsequent theorems remaining valid.

16. Specialization to ordinary curves. Theorem 10.1 retains some interest even after specialization from generalized curves to ordinary curves.

If in Theorem 10.1 the curve C^* is an ordinary curve, we readily verify, by the use of Lemma 4.1 of GC, that all the comparison curves used in the proof are also ordinary curves. Hence the conclusions of that theorem are valid under the hypothesis that C_0^* gives $J(C)$ a strong relative minimum in the class of admissible ordinary curves satisfying the end conditions. We then have the following theorem.

THEOREM 16.1. *Let the hypotheses of* §1 *be satisfied, with the modification that in* (1.7) *the word "generalized" is everywhere replaced by "ordinary". Then there is a subset* M *of* $[0, 1]$ *with measure* 1, *a non-negative constant* λ^0 *and continuous functions* $\lambda^\alpha(t, r)$ $(\alpha = 1, \cdots, m)$ *such that for the function*

$$F(y, r, \lambda) = \lambda^0 f(y, r) + \lambda^\alpha \varphi^\alpha(y, r)$$

the following conclusions hold.

(I) *There are constants* c_i $(i = 1, \cdots, \nu)$ *such that the equations*

$$F_{r^i}(y_0(t), \dot{y}_0(t), \lambda(t, \dot{y}_0(t))) = c_i + \int_0^t F_{y^i}(y_0(t), \dot{y}_0(t), \lambda(t, \dot{y}_0(t)))\, dt \tag{16.1}$$

hold for all t *in* M.

(Ia) *There are constants* $e_1, \cdots, e_p$ *such that*

$$\lambda^0 g_{y_s^i}(y_0(0), y_0(1)) + e_\mu \psi^\mu_{y_s^i}(y_0(0), y_0(1)) + (-1)^{s+1} h^i(s) = 0$$

$$(s = 0, 1;\ i = 1, \cdots, \nu),$$

where $h^i(t)$ *is an abbreviation for the right member of* (16.1).

(II) *The inequality*

$$E(y_0(t), \dot{y}_0(t), \lambda(t, \dot{y}_0(t)), r) \geqq 0 \tag{16.2}$$

holds for all t *in* M *and all* r *such that* $(y_0(t), r)$ *is admissible.*

(III) *The inequality*

$$F_{r^i r^j}(y_0(t), \dot{y}_0(t), \lambda(t, \dot{y}_0(t)))\pi^i \pi^j \geqq 0 \tag{16.3}$$

holds for all t *in* M *and all sets* $(\pi^1, \cdots, \pi^\nu)$ *which satisfy the equations*

$$\varphi^\alpha_{r^i}(y_0(t), \dot{y}_0(t))\pi^i = 0 \qquad (\alpha = 1, \cdots, m).$$

From this theorem we can readily obtain a theorem on Lipschitzian minimizing curves for Lagrange problems in non-parametric form.[16] The theorem obtained is the direct analogue of Theorem 16.1, and we shall not state it in detail. It can be regarded as the generalization to Lipschitzian functions of a recent proof[17] of the multiplier rule and Weierstrass condition without assumptions of normality. It can also be regarded as a generalization, so as to include the Weierstrass condition without assumptions of normality, of Graves' proof of the multiplier rule[18] for Lipschitzian minimizing function for Lagrange problems, save that Graves did not make the restrictive assumption (1.9). It also adds the bit of information that at each point t at which $y_0'(t)$ is continuous the multipliers $\lambda^\alpha(t, y_0'(t))$ are also continuous.

[16] The device for transforming the non-parametric problem into parametric form is set forth in several places; for example, E. J. McShane, *The Jacobi condition and the index theorem in the calculus of variations*, this Journal, vol. 5(1939), pp. 184–206; in particular, p. 188.

[17] E. J. McShane, loc. cit. (footnote 8).

[18] Loc. cit. (footnote 5).

17. **Removal of a superfluous hypothesis.** In §1 we commented that hypothesis (1.9) is a convenience rather than a necessity for our proofs. If we abandon it, all our theorems remain valid, but the proofs become somewhat tedious. In the absence of hypothesis (1.9) we first subdivide the interval $-N \leqq y^i \leqq N$, $-\mathfrak{L}(C_0^*) - 1 \leqq r^i \leqq \mathfrak{L}(C_0^*) + 1$ of (y, r)-space into subintervals so small that if one of these intervals contains a point of $R(C_0^*)$, the matrix of (1.9) can be augmented so as to be non-singular on that interval and all its neighbors. This partition of (y, r)-space generates a partition of r-space. For each t in M_0, the function $\mathfrak{M}_0[t; \Phi]$ can be represented as the sum of a finite number of functionals $\mathfrak{M}^{(k)}[t; \Phi]$, each determined by the values of Φ on just one of the intervals into which we have partitioned the interval $-\mathfrak{L}(C^*) - 1 \leqq r^i \leqq \mathfrak{L}(C_0^*) + 1$. Now all our arguments, instead of being made concerning the neighborhood $\bar{U}$ and the mean $\mathfrak{M}_0$, are based on the intervals of the partition and the means $\mathfrak{M}^{(k)}$. We are thus able to obtain all the conclusions of Theorems 10.1 without use of hypothesis (1.9). The later theorems are based on Theorem 10.1, and hence also hold without use of (1.9).

Appendix

18. **A lemma on extension of range of functions.** In this section we shall establish a lemma on the extension of range of definition of functions which we have already mentioned in preceding sections. Roughly stated, if a function on a closed set can be locally extended so as to be of class $C^{(m)}$, it can be extended to the entire space so as to be of class $C^{(m)}$. The lemma seems to have some interest in itself, so we shall prove it in a form more general than is really needed in this note.

Lemma 18.1. *Let the following hypotheses be satisfied:*

(I) *R is a closed set in $(x^1, \cdots, x^m)$-space, and $f(x)$ is a function defined and continuous on R.*

(II) *The numbers $k_1, \cdots, k_p, a_0, \cdots, a_p$ are non-negative integers such that*

$$0 = a_0 < a_1 < \cdots < a_p = m.$$

(III) *To each x_0 in R there corresponds a neighborhood $U(x_0)$ and a function $\varphi(x; x_0)$ defined and continuous on $U(x_0)$ and such that*

(a) *$\varphi(x; x_0) = f(x)$ for all x in RU;*

(b) *for each integer q ($q = 1, \cdots, p$) the function $\varphi(x; x_0)$ is of class $C^{(k_q)}$ as a function of $(x^{a_{q-1}+1}, \cdots, x^{a_q})$ wherever it is defined, the other coördinates x^i being held fixed.*

Then there exists a function $F(x)$, defined and continuous for all x, coinciding with $f(x)$ on R, and of class $C^{(k_q)}$ in the variables $(x^{a_{q-1}+1}, \cdots, x^{a_q})$ for all fixed values of the other variables.

We may suppose the neighborhoods $U(x_0)$ of hypotheses (III) to be of the form

$$x_0^i - \epsilon < x^i < x_0^i + \epsilon \qquad (i = 1, \cdots, m; \epsilon > 0).$$

Corresponding to such a neighborhood $U(x_0)$ we define $U_1(x_0)$ to be the interval

$$x_0^i - \delta < x^i < x_0^i + \delta,$$

where δ is the smaller of 1 and $\frac{1}{2}\epsilon$.

The x-space can be divided into a denumerable set of unit cubes

$$W_h : n_{h,i} \leqq x^i \leqq n_{h,i} + 1,$$

where the $n_{h,i}$ are integers. The closed set RW_h is covered by a finite number of the intervals $U_1(x_0)$. The part of W_h not contained in the finite sum of intervals is closed, and has positive distance ϵ_h from R. We cover it with a finite number of open intervals of edge less than the smaller of 1 and ϵ_h; these intervals we call "intervals of the second kind". The closures of these intervals then do not contain any points of R. Now we combine into a single sequence all the intervals $U_1(x)$ used in covering $RW_1, RW_2, \cdots$ and all the intervals of the second kind used in completing the covering of $W_1, W_2, \cdots$, and we call the resulting sequence $\delta_1, \delta_2, \cdots$. Then $\sum \delta_i$ covers the entire x-space, and each bounded set of points is contained in the sum of a finite number of intervals δ_i.

If δ_j is defined by

$$\delta_j : \alpha_j^i < x^i < \beta_j^i \qquad (i = 1, \cdots, q),$$

we define $h_j(x)$ to be the exponential of

$$-\sum_{i=1}^{n} \{(x^i - \alpha_j^i)^{-2} + (x^i - \beta_j^i)^{-2}\}$$

in δ_j, setting $h_j(x) \equiv 0$ outside δ_j. This function is of class C^∞, is positive in δ_j, and vanishes elsewhere. On each bounded set of x all but a finite number of the $h_j(x)$ vanish identically.

Next we define

$$w_j(x) = h_j(x) / \sum_{j=1}^{\infty} h_j(x).$$

These functions are of class C^∞, and

$$\sum_{j=1}^{\infty} w_j(x) \equiv 1. \tag{18.1}$$

Except on δ_j, the function $w_j(x)$ vanishes.

Each δ_j either is one of the neighborhoods $U_1(x_0)$ or else is of the second kind and therefore has no point in common with R. In the first case we define $\varphi_j(x)$ to be the function $\varphi(x; x_0)$ corresponding by hypothesis (III) to $U(x_0)$; in the second case, we define $\varphi_j(x) \equiv 0$.

We now prove that the function

$$F(x) \equiv \sum_{j=1}^{\infty} w_j(x)\varphi_j(x) \tag{18.2}$$

has the properties specified in the conclusion. Since on each bounded set the sum in (18.2) is really a finite sum and the coefficients $w_i(x)$ are of class C^∞, the continuity and differentiability properties of the F are evident. If $\bar{x}$ is in R, every δ_i which contains $\bar{x}$ is an interval $U_1(x)$, so for all such δ_i we have $\varphi_i(\bar{x}) = f(\bar{x})$, and

$$w_i(\bar{x})\varphi_i(\bar{x}) = w_i(\bar{x})f(\bar{x}). \tag{18.3}$$

For the other δ_i, which do not contain $\bar{x}$, the functions $w_i(\bar{x})$ vanish, so (18.3) is still valid. From (18.3), with (18.1) and (18.2), we at once obtain $F(\bar{x}) = f(\bar{x})$, completing the proof of the lemma.

THE UNIVERSITY OF VIRGINIA.

EXISTENCE THEOREMS FOR BOLZA PROBLEMS IN THE CALCULUS OF VARIATIONS

By E. J. McShane

This paper is the third of a sequence. The first[1] paper of the three was devoted to the development of the theory of generalized curves originated by L. C. Young, and to the establishment of existence theorems for variational problems in which the curves admitted are generalized curves. In the second paper[2] we studied the necessary conditions satisfied by generalized curves which yield strong relative minima for Bolza problems. Here we shall combine the results of the preceding two papers to obtain existence theorems in which the minimizing curve found is a curve in the ordinary sense, and not a generalized curve. The theorems thus obtained are of considerable generality.

In §§1, 3 we set forth what might be called the every-day assumptions concerning the functions and curves involved in the problem. §2 contains some remarks about the relationship between generalized and ordinary curves. In §6 the first existence theorem is stated; its proof occupies §§4, 5, 6. A generalization of this theorem is established in §7. In §8 we establish a third theorem actually somewhat less general than that of §7, but having the desirable feature that its hypotheses are stated in terms of the data of the problem, without reference to the auxiliary problem of the minimizing generalized curve. The remaining two-thirds of the paper is devoted to the proof of corollaries immediately applicable to large classes of problems, and to the study of particular examples.

1. **Assumptions concerning the functions.** The principal object of study will be a functional

$$J(C) = g(y(a), y(b)) + \mathcal{F}(C) = g(y(a), y(b)) + \int_a^b f(y(t), \dot{y}(t))\, dt,$$

defined on a class K of curves $y = y(t)$ $(a \leqq t \leqq b)$ satisfying certain equations $\varphi^\alpha(y, \dot{y}) = 0$.

We now set forth our requirements on the nature of the functions g, f, φ^α and the class K of curves.

Throughout this paper we shall assume that

(1.1) *$f(y, r)$ and $\varphi^\alpha(y, r)$ $(\alpha = 1, \cdots, m < \nu - 1)$ are defined and continuous for all[3] y in a closed set E of y-space and for all r, and are positively homogeneous*

Received November 28, 1939.

[1] E. J. McShane, *Generalized curves*, this Journal, vol. 6(1940), pp. 513–536; henceforth cited as GC.

[2] E. J. McShane, *Necessary conditions in generalized-curve problems of the calculus of variations*, this Journal, vol. 7(1940), pp. 1–27; henceforth cited as NC.

[3] As before, y is a ν-tuple $(y^1, \cdots, y^\nu)$, and likewise r.

of degree 1 *in the variables* r; *and if* y_0 *is interior to* E *and* r_0 *is a non-null vector such that*

$$\varphi^\alpha(y_0, r_0) = 0 \qquad (\alpha = 1, \cdots, m),$$

then the first-order partial derivatives of $f(y, r)$ *and* $\varphi^\alpha(y, r)$ $(\alpha = 1, \cdots, m)$ *are defined and continuous for all* (y, r) *near* (y_0, r_0).

(1.2) *If* y *is interior to* E *and* $|r| \neq 0$, *and the equations*

$$\varphi^\alpha(y, r) = 0 \tag{1.3}$$

are satisfied, the matrix

$$\| \varphi^\alpha_{r^i}(y, r) \| \qquad (\alpha = 1, \cdots, m;\ i = 1, \cdots, \nu) \tag{1.4}$$

has rank m.

2. Isomorphs of ordinary curves. We now recall that a generalized curve

$$C^*: [y(t), \mathfrak{M}[t; \Phi], M] \tag{2.1}$$

is an isomorph of an ordinary curve[4] if for almost all t in M the functional $\mathfrak{M}[t; \Phi]$ depends on the value of $\Phi(r)$ at a single point. Each such curve is the isomorph of its own track,[5] in the sense that

$$\int_a^b \mathfrak{M}[t; F(y(t), r)]\, dt = \int_a^b F(y(t), \dot{y}(t))\, dt \tag{2.2}$$

for every parametric integrand $F(y, r)$. The generalized curve C^* satisfies the differential equations (1.3) if[6] the equations

$$\int_a^b \mathfrak{M}[t; |\varphi^\alpha(y(t), r)|]\, dt = 0 \qquad (\alpha = 1, \cdots, m) \tag{2.3}$$

hold. In particular, if C^* is the isomorph of an ordinary curve, by (2.2) and (2.3) we have

$$\varphi^\alpha(y(t), \dot{y}(t)) = 0 \qquad (\alpha = 1, \cdots, m) \tag{2.4}$$

for almost all t, so that the track of C^* satisfies the differential equations (1.3) in the usual sense.

Thus if C^* is an isomorph of an ordinary curve which satisfies the equations (1.3), its track also satisfies these equations and gives the same value to the integral of f. Consequently there will be no danger of confusion if we henceforth drop the expression "isomorphs of ordinary curves" and speak simply of "ordinary curves".

3. The class of curves admitted. We shall denote by K^* the class of generalized curves in which a minimizing curve for $J(C^*)$ is sought, and by K the

[4] GC, p. 519.
[5] GC, Lemma 4.2.
[6] GC, p. 529.

subclass consisting of the ordinary curves in K^*. For simplicity, we restrict our attention to classes of curves defined by the following conditions.

(3.1) *P is a closed point-set in 2ν-dimensional space. E is a closed point-set in y-space. The class K^* is the class of all generalized curves*

$$C^*: [y(t), \mathfrak{M}[t; \Phi], M]$$

which satisfy the differential equations $\varphi^\alpha = 0$, have tracks lying in E, and have end values such that the point $(y^1(a), \cdots, y^\nu(a), y^1(b), \cdots, y^\nu(b))$ is in P.

The function $g(y_1, y_2)$ is supposed to be defined and continuous on P. No other requirement is made on g save in §§13–15, where the additional demands are specifically stated as hypotheses.

We shall assume that K^*, g, f and the φ^α are such that

(3.2) *There exists a minimizing*[7] *sequence $\{C_n^*\}$ for $\mathcal{J}(C^*)$ on K^* such that* (i) *the tracks of the C_n^* all lie in a bounded set and* (ii) *the lengths of the C_n^* are all less than a constant (independent of n).*

The next three lemmas exhibit conditions on the data of the problem which guarantee that condition (3.2) holds.

LEMMA 3.1. *If E is bounded, and*

$$f(y, r) > 0 \tag{3.3}$$

whenever y is in E and $|r| > 0$ and equations (1.3) *hold, then condition* (3.2) *is satisfied.*

Let $E \times E$ be the set of points (y_1, y_2) of 2ν-space such that y_1 and y_2 both belong to E. This set is bounded and closed; so therefore is $P \cdot [E \times E]$. On this last set the function $g(y_1, y_2)$ is continuous, hence bounded below, say by a number k. Then for curve C^* we have

$$g(y(a), y(b)) \geqq k.$$

If $\{C_n^*\}$ is a minimizing sequence, the numbers $J(C_n^*)$ are bounded above, say by a number h. Then

$$\mathcal{F}(C_n^*) = J(C_n^*) - g(y_n(a), y_n(b)) \leqq h - k.$$

Part (i) of (3.2) is evidently satisfied by the $\{C_n^*\}$. The set of (y, r) such that y is in E, $|r| = 1$ and equations (1.3) hold is bounded and closed, and on it inequality (3.3) is satisfied. Hence on this set $f(y, r)$ has a positive lower bound m. By homogeneity,

$$f(y, r) \geqq m |r|$$

[7] That is, a sequence of curves of K^* such that $J(C_n^*)$ tends to the greatest lower bound of $J(C^*)$ on K^*.

whenever y is in E and equations (1.3) hold. Hence for the curves of the sequence $\{C_n^*\}$ we have

$$\begin{aligned}\mathfrak{L}(C_n^*) &= \int_a^b \mathfrak{M}_n[t; |r|]\,dt \\ &\leqq m^{-1}\int_a^b \mathfrak{M}_n[t; f(y_n(t), r)]\,dt \\ &= m^{-1}\mathfrak{F}(C^*) \leqq (h-k)m^{-1}.\end{aligned}$$

This shows that part (ii) of (3.2) is satisfied, and completes the proof of the lemma.

For unbounded sets E we have the following lemma.

Lemma 3.2. *If*

(i) *the track of every curve of K^* has a point in common with a bounded set E_0 in y-space;*

(ii) *whenever y is in E, and $|r| \neq 0$, and the equations* (1.3) *hold, the inequality*

$$f(y, r) > 0$$

is satisfied;

(iii) *there are a positive number h and a function $\psi(y)$, defined and continuous together with its first-order partial derivatives for all y in E such that $|y| \geqq h$ and tending to ∞ as $|y| \to \infty$ in E, such that the inequality*

$$f(y, r) \geqq |\psi_{y^i}(y)r^i|$$

holds for all (y, r) with y in E and $|y| \geqq h$ which satisfy equations (1.3);

(iv) *$g(y_1, y_2)$ is bounded below;*

then condition (3.2) *is satisfied.*

Let $\{C_n^*\}$ be a minimizing sequence for $J(C^*)$. The numbers $J(C_n^*)$ have a finite upper bound. By hypothesis (iv), the numbers

$$\mathfrak{F}(C_n^*) = J(C_n^*) - g(y_n(a), y_n(b))$$

also have a finite upper bound.

Now the proof that all the C_n^* have tracks lying in a bounded set is essentially the same as for ordinary curves[8,9] if we recall Lemmas 10.1 and 2.2 of GC. The proof of Lemma 3.1 then establishes the boundedness of the lengths of the C_n^*. It is possible to extend the result slightly;[10] we may replace the inequality in (iii) by

$$f(y, r) \geqq \psi_{y^i}(y)r^i$$

[8] S. Cinquini, *Sopra l'esistenza della soluzione*, etc., Annali della R. Sc. Norm. Sup. di Pisa, ser. II, vol. 5(1936), pp. 169–190.

[9] E. J. McShane, *Some existence theorems for problems in the calculus of variations*, this Journal, vol. 4(1938), pp. 132–156; in particular, §9.

[10] Loc. cit. (footnote 9).

if we assume that the initial points of the curves of K^* lie in a bounded set, and we may replace it by

$$f(y, r) \geqq -\psi_{y^i}(y)r^i$$

if we assume that the terminal points lie in a bounded set.

A corollary of Lemma 3.2 is

LEMMA 3.3. *Condition* (3.2) *is satisfied if* (i) *the tracks of the curves of* K^* *all have points in common with a bounded set, and* (ii) *the function* $g(y_1, y_2)$ *has a finite lower bound and* (iii) *there exists a* $k > 0$ *such that the inequality*

$$f(y, r) \geqq \frac{k|r|}{|y| + 1} \tag{3.4}$$

holds for all (y, r) *which satisfy the equations* $\varphi^\alpha(y, r) = 0$.

This follows from Lemma 3.2 if we choose[11] $h = 1$ and

$$\psi(y) = k \log (|y| + 1).$$

4. **A fundamental lemma.** In §16 of NC, we defined the concept of a supporting set $S[y, \lambda^0, l]$ and of the multipliers $\lambda^\alpha(r)$ defined on such a set. These supporting sets have a close relationship to minimizing curves, as Lemma 13.2 of NC shows. For our purposes we single out a special kind of supporting set by the following definition.

(4.1) A supporting set $S_1[y, \lambda^0, l]$ with multipliers $\lambda^\alpha(r)$, has a first portion [last portion] KS_1 if there exists a closed convex set K in r-space with the following properties.

(a) For all non-null vectors r in K, (y, r) is admissible; that is, the equations $\varphi^\alpha(y, r) = 0$ are satisfied and the matrix (1.4) has rank m.

(b) On K the function $f(y, r)$ is convex.

(c) KS_1 contains at least one non-null vector.

(d) If r_0 is a non-null vector in KS_1 and r_1 is a non-null vector in $S_1 - K$, then

$$\Omega(y, r_0, r_1, \lambda) < 0, \qquad [(4.2') \; \Omega(y, r_0, r_1, \lambda) > 0], \tag{4.2}$$

where as usual we define

$$\Omega(y, r_0, r_1, \lambda) \equiv r_0^i F_{y^i}(y, r_1, \lambda(r_1)) - r_1^i F_{y^i}(y, r_0, \lambda(r_0)), \tag{4.3}$$

$$F(y, r, \lambda) = \lambda^0 f(y, r) + \lambda^\alpha \varphi^\alpha(y, r). \tag{4.4}$$

The following lemma is the kernel of this paper.

LEMMA 4.1. *Let* C_0^* : $[y_0(t), \mathfrak{M}_0[t; \Phi], M_0]$ *be a generalized curve in standard representation. Let* $\alpha \leqq t \leqq \beta$ *define an arc of* C_0^* *which satisfies conclusions* (I)

[11] Loc. cit. (footnotes 8, 9); also E. J. McShane, *Existence theorems for ordinary problems*, etc., Annali della R. Sc. Norm. Sup. di Pisa, ser. II, vol. 3(1934), pp. 287–315, in particular p. 303.

and (II) *of Theorem* 10.1 *of* NC *with multipliers* λ^0, $\lambda^a(t, r)$. *Let* M_1 *be the subset of* M_0 *consisting of those values of* t *such that* $S(t)$ *has a first portion or a last portion. Then for almost all* t *in* $M_1 \cdot [\alpha, \beta]$ *all vectors* r *carried at* $y_0(t)$ *are in the first portion of* $S(t)$ *if it has a first portion, and in the last portion of* $S(t)$ *if it has a last portion.*

For compactness, in the proof of this lemma we shall write $\Omega(r', r'')$ instead of $\Omega(y_0(t), r', r'', \lambda)$.

We may suppose that the conclusion of Theorem 14.1 of GC holds for all t in M_0; this can be brought about by the rejection of a set of measure zero. Let t be a point of M_0, and denote by $Q(r)$ the set of vectors carried by C_0^* at $y_0(t)$. Suppose to be specific that the supporting set $S(t)$ has a last portion; if instead it has a first portion, the discussion requires only trivial modification.

If the convex set K consists of a half-line, so that $S(t)$ has a "last direction", the proof is quite easy. Let K consist of the non-negative multiples of a vector r_0, which we may suppose to have length

$$(4.5) \qquad | r_0 | = \mathfrak{L}(C^*).$$

The function $\Omega(r_0, r)$ is non-negative for r in $S(t)$, and in particular for all vectors r carried at $y_0(t)$. By Theorem 14.1 of NC, the equation

$$(4.6) \qquad \mathfrak{M}_0[t; \Omega(r_0, r)] = 0$$

holds. Hence by Lemma 10.3 of GC we must have

$$(4.7) \qquad \Omega(r_0, r) = 0$$

for all r in $Q(t)$. But by (4.1), this can hold only if every vector r in $Q(t)$ is also in K, since $S(t)$ contains $Q(t)$.

The general case requires a more intricate reasoning. As a preliminary, we establish a formula concerning iterated means. If $\Phi(r', r'')$ is a continuous function of the 2ν variables (r', r''), then

$$(4.8) \qquad \mathfrak{M}_0'[t; \mathfrak{M}_0''[t; \Phi(r', r'')]] = \mathfrak{M}_0''[t; \mathfrak{M}_0'[t; \Phi(r', r'')]].$$

Here the left member denotes the number obtained by computing the mean of $\Phi(r', r'')$ for fixed r', and then computing the mean of the (continuous) function of r' thus obtained; the right member is analogously defined. The formula is evidently true if $\Phi(r', r'')$ is the product of a continuous function of r' and a continuous function of r'', hence it is valid if $\Phi(r', r'')$ is a polynomial in $r^{i\prime}$ and $r^{j\prime\prime}$. If $\Phi(r', r'')$ is continuous, it can be approximated uniformly by polynomials on the set

$$(4.9) \qquad | r' | \leqq \mathfrak{L}(C_0^*), \qquad | r'' | \leqq \mathfrak{L}(C_0^*),$$

and the means in (4.8) depend only on the values of Φ for such r' and r''. Hence (4.8) is valid.

Let us define

(4.10) *$d(r)$ is the distance from r to the closed set $KS(t)$,*

and

$$\psi_n(r) = \max\,[0, 1 - n\,d(r)]. \tag{4.11}$$

Thus $\psi_n(r)$ is 1 if r is in $KS(t)$ and is 0 if r has distance greater than n^{-1} from $KS(t)$. For the moment we define

$$\Phi(r', r'') = \psi_n(r')\psi_n(r'')\Omega(r', r'').$$

By the definition of Ω we find that

$$\Phi(r', r'') = -\Phi(r'', r'). \tag{4.12}$$

If we combine this with (4.8), and interchange the symbols r' and r'', we obtain

$$\begin{aligned} \mathfrak{M}_0'[t; \mathfrak{M}_0''[t; \Phi(r', r'')]] &= \mathfrak{M}_0''[t; \mathfrak{M}_0'[t; \Phi(r', r'')]] \\ &= -\mathfrak{M}_0''[t; \mathfrak{M}_0'[t; \Phi(r'', r')]] \\ &= -\mathfrak{M}_0'[t; \mathfrak{M}_0''[t; \Phi(r', r'')]], \end{aligned} \tag{4.13}$$

so that (returning to the original notation) we get

$$\mathfrak{M}_0'[t; \mathfrak{M}_0''[t; \psi_n(r')\psi_n(r'')\Omega(r', r'')]] = 0. \tag{4.14}$$

Suppose now that the lemma is false. Then some vector r_0 not in K is carried by C_0^* at $y_0(t)$. If r' is in $KS(t)$, the product

$$\{1 - \psi_1(r'')\}\Omega(r', r'') \tag{4.15}$$

is non-negative for all r'' in $Q(t)$; for if r'' is in $KQ(t)$ the first factor is zero, and if r'' is in $Q(t) - K$ both factors are positive. In particular, the expression (4.15) is positive for $r'' = r_0$. So by Lemma 10.3 of GC we have

$$\mathfrak{M}_0''[t; \{1 - \psi_1(r'')\}\Omega(r', r'')] > 0 \qquad (r' \text{ in } KS(t)). \tag{4.16}$$

The left member of (4.16) is continuous and positive on $KS(t)$, so its greatest lower bound on $KS(t)$ is a positive number 2ϵ. Hence

(4.17) *If r_1' is in $KS(t)$, then*

$$\mathfrak{M}_0''[t; \{1 - \psi_1(r'')\}\Omega(r_1', r'')] \geqq 2\epsilon.$$

Since $\Omega(r', r'')$ is continuous, there is a positive number δ such that if r_1, r_2 and r'' are vectors of length at most $\mathfrak{L}(C_0^*)$ and $|r_1 - r_2| \leqq \delta$, the inequality

$$|\Omega(r_1, r'') - \Omega(r_2, r'')| < \epsilon$$

holds. Let U be the δ-neighborhood of the set $KS(t)$. If r' is in the closure $\bar{U}$ of U, there is a vector r_1' of $KS(t)$ such that $|r' - r_1'| \leqq \delta$. Recalling that

$0 \leqq \psi_n(r'') \leqq \psi_1(r'') \leqq 1$ for all n and all r'', and recalling also that $\mathfrak{M}_0[t; 1] = 1$ and that $\Omega(r_1', r'')$ is non-negative for all vectors r'' carried which are not in $KS(t)$, we obtain

$$\begin{aligned} &\mathfrak{M}_0''[t; \{1 - \psi_n(r'')\}\Omega(r', r'')] \\ &\qquad \geqq \mathfrak{M}_0''[t; \{1 - \psi_n(r'')\}\{\Omega(r_1', r'') - \epsilon\}] \\ &\qquad \geqq \mathfrak{M}_0''[t; \{1 - \psi_n(r'')\}\Omega(r_1', r'')] - \epsilon \\ &\qquad \geqq \mathfrak{M}_0''[t; \{1 - \psi_1(r'')\}\Omega(r_1', r'')] - \epsilon \geqq \epsilon > 0. \end{aligned} \tag{4.18}$$

Consequently, if n exceeds $1/\delta$, the inequality

$$\psi_n(r')\mathfrak{M}_0''[t; \{1 - \psi_n(r'')\}\Omega(r', r'')] \geqq 0 \tag{4.19}$$

holds for all r' in $Q(t)$; for if r' is in $\bar{U}$ the first factor is non-negative and the second is positive by (4.18), while if r' is not in $\bar{U}$ the first factor vanishes.

By Theorem 14.1 of NC, we have

$$\mathfrak{M}_0''[t; \Omega(r', r'')] = 0 \tag{4.20}$$

for every r' in $S(t)$, and in particular for every r' in $Q(t)$. Combining (4.11), (4.18), (4.19) and (4.20), we find that the function

$$\psi_n(r')\mathfrak{M}_0''[t; \psi_n(r'')\Omega(r', r'')] \tag{4.21}$$

is non-positive for all r' in $Q(t)$ and is negative for r' in $KS(t)$. Since equation (4.14) holds, from Lemma 10.3 of GC we see that no vector of $KS(t)$ is carried; that is, $KQ(t)$ is empty.

Choose now an arbitrary vector r' of $KS(t)$. By (4.1), the function $\Omega(r', r'')$ is positive for all r'' in $Q(t) - K$; that is, for all r'' in $Q(t)$. By Lemma 10.3 of GC, we have

$$\mathfrak{M}_0''[t; \Omega(r', r'')] > 0. \tag{4.22}$$

But this contradicts (4.20), and our lemma is established.

5. **Another lemma.** To establish our principal theorems we shall need one more lemma.

LEMMA 5.1. *Let t be a point of M_0. If all vectors r carried by C_0^* at $y_0(t)$ are contained in a closed convex set K, and $(y_0(t), r)$ is admissible for each non-null vector r in K, and $f(y_0(t), r)$ is convex on K, then the relations*

$$f(y_0(t), \dot{y}_0(t)) \leqq \mathfrak{M}_0[t; f(y_0(t), r)] \tag{5.1}$$

and

$$\varphi^\alpha(y_0(t), \dot{y}_0(t)) = 0 \qquad (\alpha = 1, \cdots, m) \tag{5.2}$$

are satisfied.

Since K is convex, and $\mathfrak{M}_0$ is a linear mean, and $f(y_0(t), r)$ is convex on K, by Jensen's inequality[12] the point

$$(5.3) \qquad y_0'(t) \equiv (\mathfrak{M}_0[t; r^1], \cdots, \mathfrak{M}_0[t; r^r])$$

is in K, and

$$(5.4) \qquad f(y_0(t), \mathfrak{M}_0[t; r^1], \cdots, \mathfrak{M}_0[t; r^r]) \leqq \mathfrak{M}_0[t; f(y_0(t), r)].$$

Since $y_0'(t)$ is in K, either $|y_0'(t)| = 0$ or else $(y(t), y_0'(t))$ is admissible, and in either case equations (5.2) are satisfied. Inequality (5.4), with (5.3), yields (5.1), and the proof of the lemma is complete.

6. **First existence theorem.** We are now ready to establish the first of our principal theorems.

THEOREM 6.1. *Let the set E consist of the entire y-space, and let conditions* (1.1), (1.2), (3.1) *and* (3.2) *be satisfied. Let*

$$C_0^* : [y_0(t), \mathfrak{M}_0[t; \Phi], M_0]$$

be the standard representation of a generalized curve which minimizes[13] *the functional $J(C^*)$ in the class K^*; for this curve there then exists at least one set of multipliers λ^0, $\lambda^\alpha(t, r)$ with which conclusions* (I) *and* (II) *of Theorem* 10.1 *of* NC *hold. Let it be true that these multipliers can be so chosen that for almost all t in M_0 the supporting set $S(t)$ has a first portion or a last portion.*

Then the track of C_0^* is itself a minimizing curve for $J(C^*)$ in the class K^*.

If the hypotheses of Theorem 6.1 are satisfied, so are those of Lemma 4.1, the set M_1 constituting almost all of M_0. By Lemma 4.1, the hypotheses of Lemma 5.1 are satisfied for almost all t in M_0.

Hence (5.2) holds for almost all t, and the track of C_0^* satisfies the differential equations $\varphi^\alpha = 0$. It has the same ends as C_0^*, so it belongs to the class K^*. Since C_0^* minimizes $J(C^*)$ on K^*, we have

$$(6.1) \qquad g(y_0(a), y_0(b)) + \int_0^1 f(y_0, \dot{y}_0)\, dt \geqq g(y_0(a), y_0(b)) + \mathcal{F}(C_0^*).$$

But integration of both members of (5.1) yields

$$(6.2) \qquad \int_0^1 f(y_0, \dot{y}_0)\, dt \leqq \mathcal{F}(C_0^*).$$

This, with (6.1), implies that the members of (6.1) are equal, so that the track of C_0^* is a minimizing curve for $J(C^*)$ in the class K^*. The proof of Theorem 6.1 is complete.

It is interesting to observe that we have incidentally proved the following corollary.

[12] E. J. McShane, *Jensen's inequality*, Bull. Amer. Math. Soc., vol. 43(1937), pp. 521–527.

[13] Such a curve exists, by Theorem 8.2 of GC.

THEOREM 6.2. *Under the hypotheses of Theorem* 6.1, *the set of t for which $S(t)$ has both a first portion and a last portion without common points is a set of measure zero.*

By Lemma 4.1, for almost all such t the set $Q(t)$ is not empty and is entirely contained in $KS(t)$, where $KS(t)$ is either the first portion or the last portion. But then $Q(t)$ is contained in each of two disjoint sets, and this is impossible.

7. **Second existence theorem.** Theorem 6.1 admits of a ready generalization to more general classes K^*, in which the set E is not required to be the entire y-space.

THEOREM 7.1. *Let conditions* (1.1), (1.2), (3.1) *and* (3.2) *be satisfied. Let*

$$C_0^*: [y_0(t), \mathfrak{M}_0[t; \Phi], M_0]$$

be the standard representation of a generalized curve which minimizes $J(C^*)$ on K^*. For each t_0 in M_0, let one or the other of the following hypotheses be satisfied.

(I) *The solutions of the equations $\varphi^\alpha(y_0(t_0), r) = 0$ $(\alpha = 1, \cdots, m)$ form a convex set in r-space, and on this set the function $f(y_0(t_0), r)$ is convex.*

(II) *The number t_0 belongs to an interval $\alpha \leqq t \leqq \beta$ $(\alpha < \beta)$ which defines an arc of C_0^* whose track is interior to E. On this arc* (I) *and* (II) *of the multiplier rule of Theorem* 10.1 *of* NC *are satisfied with multipliers λ^0, $\lambda^\alpha(t, r)$ with the property that for almost all t in (α, β), the set $S(t)$ has a first portion or a last portion.*

Then the track of C_0^* is itself a minimizing curve for $J(C^*)$ in the class K^*.

Let us suppose that for all t in M_0 the functions $\dot{y}_0^i$, $\varphi^\alpha(y_0, \dot{y}_0)$ and $\mathfrak{M}_0[t; f(y_0(t), r)]$ are the derivatives of their indefinite integrals; this can be brought about by discarding a set of measure zero. Let t_0 be a point of M_0. If hypothesis (I) is satisfied at t_0, by Lemma 5.1 we find

$$f(y_0(t_0), \dot{y}_0(t_0)) \leqq \mathfrak{M}_0[t_0; f(y_0(t_0), r)] \tag{7.1}$$

and

$$\varphi^\alpha(y_0(t_0), \dot{y}_0(t_0)) = 0 \qquad (\alpha = 1, \cdots, m). \tag{7.2}$$

Suppose then that hypothesis (II) holds at t_0. By Lemmas 4.1 and 5.1 relations (5.1) and (5.2) hold at all points t of the interval (α, β) save those belonging to a subset of measure zero. The set of all points t_0 at which hypothesis (II) holds can be covered by a denumerable set of the corresponding intervals (α, β). Rejecting from each of these the subset (of measure zero) on which (5.1) and (5.2) do not both hold, we find that for almost all t_0 at which hypothesis (II) holds relations (7.1) and (7.2) are also satisfied.

The proof of Theorem 6.1 can now be repeated to complete the proof of Theorem 7.1.

8. **Third existence theorem.** The two existence theorems already established involve hypotheses on the multipliers λ^0, $\lambda^\alpha(t, r)$. It is desirable to have a theorem involving only the given data of the problem. In this section we establish such a theorem. For brevity, we assign the name "ordinary points" to the points at which our methods are certain to apply; specifically, we adopt the following definition.

(8.1) *A point y_0 of E is ordinary if it satisfies either one of the two following conditions.*

(a) *The aggregate of solutions r of the equations $\varphi^\alpha(y_0, r) = 0$ $(\alpha = 1, \cdots, m)$ is a convex set, and on it the function $f(y_0, r)$ is convex.*

(b) *The point y_0 is interior to E, and every supporting set $S[y_0, \lambda^0, l]$ has either a first portion or a last portion.*

Save for the requirement of interiority, (b) is weaker than (a); for if (a) is satisfied, every supporting set is contained in the convex set of solutions of $\varphi^\alpha = 0$, and is therefore its own first (and last) portion.

Our third existence theorem is

THEOREM 8.1. *Let conditions* (1.1), (1.2), (3.1) *and* (3.2) *be satisfied. Let every point of E be ordinary. Then if the class K^* is not empty, it contains an ordinary curve C_0 which minimizes $J(C^*)$ on the class K^*, and a fortiori on the subclass K.*

The hypotheses of Theorem 8.1 of GC are satisfied, so there exists a minimizing curve C_0^* for $J(C^*)$ in the class K^*. We suppose that

$$C_0^* : [y_0(t), \mathfrak{M}_0[t; \Phi], M_0]$$

is the standard representation of C_0^*. Let t_0 be an arbitrary point of the interval $[0, 1]$. If $y_0(t_0)$ is a boundary point of E, condition (8.1a) must hold; then hypothesis (I) of Theorem 7.1 is satisfied. If $y_0(t_0)$ is interior to E, every point in a neighborhood U of $y_0(t_0)$ is also interior to E. Hence at all such points condition (8.1b) holds. If $[\alpha, \beta]$ is any sufficiently small interval containing t_0, the track of the corresponding arc of C_0^* lies in U. This arc minimizes the integral of f in the class of arcs joining its ends, lying sufficiently near it and satisfying the equations $\varphi^\alpha = 0$. Hence it satisfies (I) and (II) of the multiplier rule of Theorem 10.1 of NC with multipliers λ^0, $\lambda^\alpha(t, r)$. Hypothesis (II) of Theorem 7.1 must then hold; for every supporting set at $y(t)$ $(\alpha \leqq t \leqq \beta)$, and in particular the supporting set $S(t)$, must have a first portion or a last portion.

9. **Free problems.** The theorems established in the preceding sections apply at once to problems without side conditions. For such problems we can choose $\lambda^0 = 1$; there are no other multipliers, and F is identical with f. All (y, r) with y in E are admissible.

In a previous paper[14] I have established some theorems for free problems in

[14] E. J. McShane, *Some existence theorems in the calculus of variations*, III. *Existence theorems for non-regular problems*, Trans. Amer. Math. Soc., vol. 45(1939), pp. 151–171.

parametric form, without assumptions of regularity. These theorems are obvious corollaries of Theorem 8.1.

To obtain an example covered by Theorem 8.1 (and not, to the best of my knowledge, by any previously published theorem) we consider the integrand

$$
\begin{aligned}
f(x, y, z, x', y', z') = l(k^2x'^2 + y'^2 + z'^2)^{\frac{1}{2}} \\
+ \lambda_1\Phi(x, y, z)(x'^2 + y'^2 + z'^2)^{\frac{1}{2}},
\end{aligned}
\tag{9.1}
$$

where we suppose $k > 1$. For the set E we take the entire y-space. This integrand satisfies condition (3.6), and therefore satisfies condition (3.2), if there is a positive number c such that

$$
l + \lambda_1\Phi(x, y, z) \geqq c[1 + (x^2 + y^2 + z^2)^{\frac{1}{2}}]^{-1}. \tag{9.2}
$$

We suppose this condition satisfied; furthermore, we assume

$$
l \geqq 0. \tag{9.3}
$$

If (9.2) is satisfied, f is positive for $(x', y', z') \neq (0, 0, 0)$, and for fixed (x, y, z) the equation

$$
f(x, y, z, x', y', z') = 1 \tag{9.4}
$$

defines a surface Σ_1 in (x', y', z')-space. This is a surface of revolution about the x'-axis. Consider a supporting plane of Σ_1. We may suppose, because of the rotational symmetry, that this plane meets Σ_1 at a point with $z' = 0$. It is easy to see that if this plane has more than one point of contact with Σ_1, it must have more than one point of contact with Σ_1 at which $z' = 0$. Therefore in seeking supporting sets we may set $z' = 0$.

Letting $z' = 0$ gives us essentially a plane problem, for which (in the usual notation)

$$
F_1 = (y')^{-2}f_{x'x'} = lk^2(k^2x'^2 + y'^2)^{-\frac{3}{2}} + \lambda_1\Phi(x'^2 + y'^2)^{-\frac{3}{2}}.
$$

In investigating the sign of this function we may restrict our attention to sets with $x'^2 + y'^2 = 1$; we then find that F_1 has a minimum at $(\pm 1, 0)$, and if the equation

$$
l/k + \lambda_1\Phi \geqq 0 \tag{9.5}
$$

holds, the function F_1 is positive save perhaps at $(\pm 1, 0)$. The surface Σ_1 is then strictly convex, and each supporting set of f consists of a single half-line from the origin.

If inequality (9.5) fails, the intersection of Σ_1 with the plane $z' = 0$ is a dumbbell-shaped curve, pinched in along the x'-axis. This has tangent (supporting) lines $x' = c_1 =$ constant which meet it at two places. Hence Σ_1 has supporting planes $x' =$ constant which meet Σ_1 along a circle

$$
x' = c_1, \qquad y'^2 + z'^2 = c_2^2 = \text{constant}. \tag{9.6}
$$

Let S be a supporting set defined by equations (9.6). We suppose $x'^2 + y'^2 + z'^2 = 1$, and compute the Ω-function; this is

$$(9.7)\quad \Omega(x, y, z, x', y', z', \bar{x}', \bar{y}', \bar{z}') = \lambda_1[c_1^2 + c_2^2]^{\frac{1}{2}}[(y' - \bar{y}')\Phi_y + (z' - \bar{z}')\Phi_z].$$

Suppose now that

$$(9.8)\qquad \Phi_y^2 + \Phi_z^2 > 0.$$

The point

$$(9.9)\qquad x' = c_1, \qquad y' = c_2\Phi_y(\Phi_y^2 + \Phi_z^2)^{-\frac{1}{2}}, \qquad z' = c_2\Phi_z(\Phi_y^2 + \Phi_z^2)^{-\frac{1}{2}}$$

is in the set (9.6). Since (9.3) holds and (9.5) does not, λ_1 is not zero. The constant c_2 in (9.6) can be chosen positive or negative; we choose the sign of c_2 to be the same as that of λ_1. Now the Cauchy-Schwarz inequality applied to (9.7) shows that (9.9) is the last element of the supporting set (9.6). Thus if (9.8) is satisfied, every supporting set at (x, y, z) has a last element. If we replace c_2 by $-c_2$ in (9.9), we obtain a first element. Thus from Theorem 8.1 we obtain the following result. *If A_1 and A_2 are closed point sets in (x, y, z)-space, one of which is bounded, and k, l, λ_1, and Φ are such that at each point (x, y, z) conditions* (9.2) *and* (9.3) *hold, as well as one of the conditions* (9.5), (9.8), *then the class of curves joining A_1 to A_2 contains a minimizing curve for $\int f\,dt$.*

We may for instance take $l = 1$, $k = 2$, $\lambda_1 = -1$, $\Phi = 1 - e^{-(y^2+z^2)}$.

10. **Isoperimetric problems.** As we observed in §15 of NC, the isoperimetric problem of minimizing $\mathcal{F}(C^*)$ subject to the conditions $\mathcal{G}^\alpha(C^*) = \gamma^\alpha$ ($\alpha = 1, \cdots, m$) can be reduced to a Bolza problem by introducing new variables $y^{\nu+\alpha}$, $r^{\nu+\alpha}$ ($\alpha = 1, \cdots, m$) and replacing the isoperimetric conditions by the differential equations

$$(10.1)\qquad r^{\nu+\alpha} + g^\alpha(y(t), r) = 0 \qquad (\alpha = 1, \cdots, m)$$

and the end conditions

$$y^{\nu+\alpha}(0) = 0, \qquad y^{\nu+\alpha}(1) = -\gamma^\alpha \qquad (\alpha = 1, \cdots, m).$$

In the multiplier rule these new variables can be omitted, and the multipliers can be chosen to be constants. It is therefore desirable to rephrase the definitions of supporting sets and first and last elements so as to avoid reference to the new variables.

By definition (§13 of NC), the set $S[y_0, \lambda^0, l]$ of admissible elements is a supporting set if the inequality

$$\lambda^0 f(y_0, r) - r^\gamma l^\gamma \geqq 0 \qquad (\gamma = 1, \cdots, \nu + m)$$

holds for all admissible elements (y_0, r), equality holding on $S(y_0, \lambda^0, l)$. By (10.1), this is equivalent to assuming

$$[\lambda^0 f(y_0, r) + l^{\nu+\alpha} g^\alpha(y_0, r)] - l^i r^i \geqq 0 \qquad (\alpha = 1, \cdots, m; i = 1, \cdots, \nu)$$

for all r, equality holding if $(y_0, r^1, \cdots, r^\nu, -g^1(y_0, r), \cdots, -g^m(y_0, r))$ is in S. The multipliers of the set are then $(\lambda^0, \lambda^\alpha(r)) = (\lambda^0, l^{\nu+1}, \cdots, l^{\nu+m})$. There is a one-to-one correspondence between the points of S and the points of a set S_0 in the 2ν-dimensional space of points (y, r), obtained by omission of the last m coördinates of the points of S. We shall henceforth refer to S_0 as the supporting set. Thus S_0 is a supporting set at y_0, with multipliers λ^0, $\lambda^1, \cdots, \lambda^m$, if there is a linear function l^ir^i such that $F(y, r, \lambda) \equiv \lambda^0 f + \lambda^\alpha g^\alpha$ coincides with l^ir^i on S_0 and exceeds l^ir^i for r not in S_0.

Suppose next that $S(y_0, \lambda^0, l)$ has a first portion K. If r_1 and r_2 are points of $(\nu + m)$-space both belonging to K, by the convexity of K the point $\frac{1}{2}(r_1 + r_2)$ is also in K, hence is admissible. That is,

$$\left(\tfrac{1}{2}(r_1^1 + r_2^1), \cdots, \tfrac{1}{2}(r_1^\nu + r_2^\nu), -\tfrac{1}{2}(g^1(y_0, r_1) + g^1(y_0, r_2)), \cdots, \right.$$
$$\left. -\tfrac{1}{2}(g^m(y_0, r_1) + g^m(y_0, r_2))\right)$$

is admissible, so that

$$g^\alpha(y_0, \tfrac{1}{2}(r_1 + r_2)) = \tfrac{1}{2}[g^\alpha(y_0, r_1) + g^\alpha(y_0, r_2)].$$

If we project K on $(y, r^1, \cdots, r^\nu)$-space by omitting the last m coördinates, we obtain a convex set K_0. The preceding equation then implies that the integrands $g^\alpha(y_0, r)$ are linear on K_0. Conversely, suppose that K_0 is convex and the g^α are linear on K_0. The set K of points $(y_0, r^1, \cdots, r^\nu, -g^1(y_0, r), \cdots, -g^m(y_0, r))$ with (y_0, r) in K_0 is the linear image of a convex set, and is therefore convex.

The new variables $y^{\nu+\alpha}$, $r^{\nu+\alpha}$ vanish from the definition of the Ω-function; so for isoperimetric problems definition (4.1) is equivalent to the following.

(10.2) *A supporting set S_0 with multipliers λ^α has a first portion [last portion] K_0S_0 if there exists a convex set K_0 in $(r^1, \cdots, r^\nu)$-space with the following properties.*

(a) *On K_0 the functions $g^\alpha(y, r)$ are linear and the function $f(y, r)$ is convex.*

(b) *K_0S_0 contains at least one non-null vector.*

(c) *If r_0 is a non-null vector in K_0S_0, and r_1 is in $S_0 - K_0$, then*

$$\Omega(y, r_0, r_1, \lambda) < 0 \qquad [\Omega(y, r_0, r_1, \lambda) > 0].$$

Consider, for example, the problem of minimizing

$$\mathcal{F}(C) = \int_C (k^2x'^2 + y'^2 + z'^2)^{\frac{1}{2}}\, dt \qquad (k > 1) \tag{10.3}$$

in the class of curves C joining two fixed point sets A_1, A_2 (one of them being bounded) and giving a constant value γ to the integral

$$\mathcal{G}(C) = \int_C \Phi(x, y, z)(x'^2 + y'^2 + z'^2)^{\frac{1}{2}}\, dt. \tag{10.4}$$

We suppose that Φ_y and Φ_z never vanish simultaneously, so that condition (9.8) holds for all (x, y, z).

On every minimizing sequence $\mathcal{F}(C)$ is bounded, so the length $\mathcal{L}(C)$ is bounded, and hypothesis (3.2) holds. We must show that every point of (x, y, z)-space is ordinary. Where Φ vanishes condition (8.1a) is readily seen to hold; therefore we need consider only points at which $\Phi \neq 0$. Suppose then that S is a supporting set at (x, y, z), and let $\lambda_0 \geqq 0$, λ_1 be the multipliers for S. Define

$$(10.5)\qquad \begin{aligned} F(x, y, z, x', y', z', \lambda_0, \lambda_1) = \lambda_0(k^2x'^2 + y'^2 + z'^2)^{\frac{1}{2}} \\ + \lambda_1\Phi(x'^2 + y'^2 + z'^2)^{\frac{1}{2}}, \end{aligned}$$

and let l_1, l_2, l_3 be the respective constant values of $F_{x'}$, $F_{y'}$, $F_{z'}$ over S. By definition of supporting sets (NC, §13),

$$F(x, y, z, 0, 1, 0, \lambda_0, \lambda_1) \geqq l_2 \cdot 1, \qquad F(x, y, z, 0, -1, 0, \lambda_0, \lambda_1) \geqq l_2 \cdot (-1),$$

whence by adding we find

$$(10.6)\qquad \lambda_0 + \lambda_1\Phi(x, y, z) \geqq 0.$$

If equality holds in (10.6), then F vanishes if $x' = 0$ and is positive elsewhere. In this case the plane $x' = 0$ is a supporting set. If the left member of (10.6) is positive, but (9.5) fails (with $l = \lambda_0$), there are supporting sets which are circular cones with vertex at $(0, 0, 0)$ and axis along the x'-axis; the supporting set $x' = 0$ corresponding to equality in (10.6) may be considered as belonging to this type. As shown in §9, since inequality (9.8) holds, these sets have a first portion and a last portion. Also, as in §9, if inequality (9.5) holds, each supporting set consists of a single half-line issuing from the origin. Thus in all cases every supporting set has a first portion and a last portion, and so every point of (x, y, z)-space is ordinary. Therefore, by Theorem 8.1, if there exist curves C joining A_1 to A_2 and having $\mathcal{G}(C) = \gamma$, this class of curves contains a minimizing curve for $\mathcal{F}(C)$.

11. **An existence theorem for isoperimetric problems.** Many interesting isoperimetric problems involve integrals whose integrands are closely related. We now prove a corollary of Theorem 8.1 applying to such problems. Specifically, we are to minimize an integral $\mathcal{F}(C)$ in the class of curves for which certain integrals $\mathcal{G}^1(C), \cdots, \mathcal{G}^m(C)$ assume assigned values. The relationship between the integrands is that there are a function $\varphi(y, r)$ and a set of functions $h^0(y), \cdots, h^m(y)$ such that

$$(11.1)\qquad \begin{aligned} f(y, r) &= h^0(y)\varphi(y, r), \\ g^\alpha(y, r) &= h^\alpha(y)\varphi(y, r) \qquad (\alpha = 1, \cdots, m). \end{aligned}$$

For such systems we shall prove the following theorem.

THEOREM 11.1. *Let the following hypotheses be satisfied.*

(I) *The function $\varphi(y, r)$ is defined and satisfies condition* (1.1) *for all y and r, and if $|r_1| \neq 0$ the inequality*

$$(11.2)\qquad \mathcal{E}(y, r_1, r_2) = \varphi(y, r_2) - r_2^i\varphi_{r^i}(y, r_1) \geqq 0$$

holds.

(II) *The inequality*

$$\varphi(y, r) > 0 \tag{11.3}$$

holds whenever $|r| > 0$.

(III) *The functions* $h^\gamma(y)$ $(\gamma = 0, 1, \cdots, m \leqq \nu)$ *are defined and of class* C′ *for all* y, *and the matrix*

$$\left\| \begin{matrix} h^\gamma(y) \\ h^\gamma_{y^i}(y) \end{matrix} \right\| \qquad (\gamma = 0, 1, \cdots, m;\ i = 1, \cdots, \nu) \tag{11.4}$$

has rank $m + 1$ *for all* y.

(IV) K *is the class of all rectifiable curves satisfying a set of end conditions such as described in* (3.1).

(V) *The numbers* $\gamma^1, \cdots, \gamma^m$ *are constants such that the equations*

$$\mathcal{G}^\alpha(C) = \int_a^b h^\alpha(y)\varphi(y, \dot{y}')\, dt = \gamma^\alpha \qquad (\alpha = 1, \cdots, m) \tag{11.5}$$

are satisfied by at least one curve C: $y = y(t)$ $(a \leqq t \leqq b)$ *of the class* K.

(VI) *Condition* (3.2) *is satisfied.*

Then the subclass of K *consisting of those curves of* K *which satisfy equations* (11.5) *contains a curve* C *on which*

$$\mathcal{F}(C) = \int_a^b h^0(y)\varphi(y, \dot{y})\, dt$$

assumes its least value.

This theorem will follow at once from Theorem 8.1 if we can show that every point y is ordinary.

Hypothesis (I) implies that $\varphi(y, r)$ is a convex function of r for each fixed y. Let us denote by $K(r_0)$ the set of all r for which the equation

$$\varphi(y, r) - r^i\varphi_{r^i}(y, r_0) = 0 \tag{11.6}$$

holds. This is the set of r on which the convex function $\varphi(y, r)$ coincides with one of its supporting planes, so it is a convex set.

Let S be a supporting set at y, with multipliers $\lambda^0 \geqq 0$, λ^α, not all zero. In accordance with our usual notation we write

$$F(y, r, \lambda) = \lambda^\gamma h^\gamma(y)\varphi(y, r).$$

If r, $\bar{r}$ are non-null vectors of S, each derivative F_{r^i} has the same value at r and at $\bar{r}$. From this, on multiplication by $\bar{r}^i$, we obtain

$$\lambda^\gamma h^\gamma(y)\{\varphi(y, \bar{r}) - \bar{r}^i\varphi_{r^i}(y, r)\} = 0. \tag{11.7}$$

Suppose first that the factor $\lambda^\gamma h^\gamma(y)$ is different from zero. Then the second factor must vanish, so that every non-null vector r of S is in $K(\bar{r})$. On $K(\bar{r})$ the function $\varphi(y, r)$ coincides with a linear function by (11.6), so S is its own first and last portion.

Suppose next that

$$\lambda^\gamma h^\gamma(y) = 0. \tag{11.8}$$

Then $F(y, r)$ is identically zero, and S is the whole r-space. We calculate (remembering (11.8))

$$\Omega(y, r, \bar{r}, \lambda) = r^i\lambda^\gamma h^\gamma_{y^i}(y)\varphi(y, \bar{r}) - \bar{r}^i\lambda^\gamma h^\gamma_{y^i}(y)\varphi(y, r). \tag{11.9}$$

Denote by m the maximum value of the function

$$r^iH_i, \qquad H_i \equiv \lambda^\gamma h^\gamma_{y^i}(y), \tag{11.10}$$

on the bounded closed set of points r such that $\varphi(y, r) = 1$, and let this maximum be attained at r_0. The numbers H_i cannot all vanish; if they did, equations (11.8) and (11.10), with hypothesis (III), would imply $\lambda^0 = \cdots = \lambda^m = 0$, and this is impossible. It follows that the linear function r^iH_i can assume both positive and negative values on the surface $\varphi(y, r) = 1$, so that m is positive.

By the homogeneity of $\varphi(y, r)$ and the definition of m, the inequality

$$r^iH_i - m\varphi(y, r) \leqq 0 \tag{11.11}$$

holds for all r, equality holding at r_0. Hence all the partial derivatives of the left member of (11.11) must vanish at r_0, and we have

$$H_i = m\varphi_{r^i}(y, r_0) \qquad (i = 1, \cdots, \nu). \tag{11.12}$$

By virtue of (11.10) and (11.12), equation (11.9) yields

$$\begin{aligned}\Omega(y, r_0, \bar{r}, \lambda) &= m\varphi_{r^i}(y, r_0)\,[r_0^i\varphi(y, \bar{r}) - \bar{r}^i\varphi(y, r_0)]\\ &= m\varphi(y, r_0)[\varphi(y, \bar{r}) - \bar{r}^i\varphi_{r^i}(y, r_0)].\end{aligned} \tag{11.13}$$

Recalling that m is positive and $\varphi(y, r_0)$ is 1, we see that this last expression is positive unless $\bar{r}$ belongs to the convex set $K(r_0)$ on which $\varphi(y, r)$ (and with it $f(y, r)$ and all the $g^\alpha(y, r)$) is linear. That is, $K(r_0)$ is the last portion of S.

We have now shown that every point y is ordinary, and the proof of Theorem 11.1 is complete.

For an example, let K be the class of rectifiable curves joining two fixed points P_1, P_2 of the (x, y)-plane and having assigned moments about the coördinate axes:

$$\int_a^b x(t)(x'^2 + y'^2)^{\frac{1}{2}}\,dt = \mu_1, \qquad \int_a^b y(t)(x'^2 + y'^2)^{\frac{1}{2}}\,dt = \mu_2.$$

It is not difficult to show that this class is not empty. In the class K we seek a curve of least length. We take

$$\varphi = (x'^2 + y'^2)^{\frac{1}{2}}, \qquad h^0 = 1, h^1 = x, h^2 = y.$$

The matrix (11.4) is

$$\begin{vmatrix} 1 & x & y \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{vmatrix},$$

which is non-singular. The other hypotheses of Theorem 11.1 are readily verified, so the minimizing curve exists.

It is interesting to observe that the functions f, g^α enter symmetrically into all the hypotheses of Theorem 11.1 except hypothesis (VI), which concerns the existence of a convergent minimizing sequence.

12. **Problems with a single side-integral.** In a recent paper[15] (henceforth referred to as IP) I have established existence theorems of some generality for isoperimetric problems involving a single side-integral. We are now in a position to generalize these theorems in several respects, and at the same time to shorten their proofs. For brevity, we shall content ourselves with indicating the changes which are to be made in the paper cited, retaining the notation of that paper.

In §2 of IP the definition of ordinary point (2.1) is to be replaced by the definition (8.1) of the present paper. The hypotheses of Theorem 2 of IP guarantee the existence of a minimizing generalized curve,

$$C_0^* : [z_0(t), \mathfrak{M}_0[t; \Phi], M_0], \tag{12.1}$$

in the class of generalized curves satisfying the end conditions and giving $\mathfrak{G}(C^*)$ the value l. We suppose that (12.1) is the standard representation of C_0^*, and for $0 \leqq t \leqq 1$ we define

$$\begin{aligned} \varphi(t) &= \int_0^t \mathfrak{M}_0[t; F(z_0(t), r)]\, dt, \\ \gamma(t) &= \int_0^t \mathfrak{M}_0[t; G(z_0(t), r)]\, dt, \\ \varphi_0(t) &= \int_0^t F(z_0, \dot z_0)\, dt, \\ \gamma_0(t) &= \int_0^t G(z_0, \dot z_0)\, dt. \end{aligned} \tag{12.2}$$

The set E is the subset of $[0, 1]$ on which all four integrals in (12.2) have derivatives equal to their respective integrands.

We abandon §3 of IP. Lemma 1 of §4 is merely Jensen's inequality, since the hypothesis is that $aF(z, r) + bG(z, r)$ is a convex function of r. As a substitute for Lemma 2 of IP we prove

LEMMA 12.1. *For almost all t_0 such that $z(t_0)$ is an ordinary point the relations*

$$\varphi'(t_0) \geqq \varphi_0'(t_0), \qquad \gamma'(t_0) = \gamma_0'(t_0) \tag{12.3}$$

are satisfied.

[15] E. J. McShane, *Some existence theorems in the calculus of variations*, V. *The isoperimetric problem in parametric form*, Trans. Amer. Math. Soc., vol. 45(1939), pp. 197–216.

If $z_0(t_0)$ is a boundary point of the set (called S in IP, called E in this paper) and is ordinary, then at $z(t_0)$ the function $F(z, r)$ is convex and $G(z, r)$ is linear. In this case (12.3) follows from Lemma 4.1.

Consider now any arc $z = z_0(t)$, $\alpha \leqq t \leqq \beta$, of the minimizing curve interior to the set S. On this arc the multiplier rule of Theorem 13.1 of NC is satisfied with multipliers $\lambda^0 \geqq 0$, λ^1, not both zero. By Lemma 4.1, for almost all points t_0 at which the supporting set $S(t)$ has a first or last portion $KS(t)$ (in particular, for almost all t_0 such that $z_0(t_0)$ is ordinary), the set of all vectors carried at $z_0(t_0)$ is contained in that first or last portion. By definition, the function $f(z_0(t_0), r)$ is convex and the function $g(z_0(t_0), r)$ is linear on K; so for almost all such t_0 we can again apply Jensen's inequality to establish (12.3).

We can choose a denumerable set of intervals $[\alpha_i, \beta_i]$ such that the sum of the arcs $z = z_0(t)$ ($\alpha_i \leqq t \leqq \beta_i$) covers the part of the track of C_0^* interior to the set called S in IP. For almost all t in each $[\alpha_i, \beta_i]$ such that $z_0(t)$ is ordinary the relations (12.4) hold; hence these relations hold for almost all t such that $z_0(t)$ is ordinary and interior to the set. As the relations have been established for ordinary boundary points, the proof of Lemma 12.1 is complete.

The remainder of the proof of Theorem 2 of IP, through §5, requires no change, save the remark that μ_0 and μ are to be regarded merely as two symbols for $\mathfrak{F}(C_0^*)$, i.e., the greatest lower bound of $\mathfrak{F}(C^*)$ on the class of curves allowed.

Theorem 3 of IP can also be generalized; in fact, in (7.3) of IP we can define $M_1(z)$ to be the g.l.b. of $M_1(z, A)$ for all sets A which are supporting sets at z and have neither first nor last portion. We omit the proof.

Theorem 4 of IP is already contained in the preceding theorems; for our present supporting sets are by definition "$\mathfrak{E}$-admissible".

13. **Mayer problems.** For problems of Mayer, which are Bolza problems with $f \equiv 0$, Graves[16] has established existence theorems (both for the parametric and non-parametric forms) which at first glance would not seem related to the theorems of this paper. We shall now see that Theorem 6.1 has a corollary which is closely related to Graves' theorem and overlaps it considerably, though neither theorem includes the other.

We shall consider problems in which the set E is the entire y-space and one particular minor of the matrix $\| \varphi^\alpha_{r^i} \|$ is non-singular. In such problems we can solve the equations $\varphi^\alpha = 0$ for m of the variables r^i (say $r^{n+1}, \cdots, r^\nu$, $n = \nu - m$) in terms of the other variables, $r^1, \cdots, r^n, y^1, \cdots, y^\nu$. If we designate $r^{n+\alpha}$ by ρ^α ($\alpha = 1, \cdots, m$) and correspondingly designate $y^{n+\alpha}$ by z^α, the equations $\varphi^\alpha(y, r) = 0$ take the form

$$h^\alpha(y, z, r) - \rho^\alpha = 0 \qquad (\alpha = 1, \cdots, m). \tag{13.1}$$

The distinctive assumptions on these functions which are used in establishing our corollary are the following.

[16] L. M. Graves, *The existence of an extremum in problems of Mayer*, Trans. Amer. Math. Soc., vol. 39(1936), pp. 456–471. In this paper there are references to earlier papers by Manià concerning the problem of Mayer.

(13.2) *The end conditions on the curves of the class K do not involve the final values of the* $z^\alpha(t)$.

In the terminology of §3, the set P of admitted end values (y_1, z_1, y_2, z_2) is such that if (y_1, z_1, y_2, z_2) is in P, so is $(y_1, z_1, y_2, \bar{z}_2)$ for all sets of numbers $\bar{z}_2$.

(13.3) *For each point in P, all the partial derivatives* $\partial g/\partial z_2^\alpha$ *are defined, continuous and non-negative, and not all of them are zero.*

(13.4) *For all* (y, z, r) *the partial derivatives*

$$h_{z\beta}^\alpha(y, z, r) \qquad (\alpha \neq \beta;\ \alpha, \beta = 1, \cdots, m)$$

are non-negative.

(13.5) *For all* (y, z, r) *with* $|r| > 0$ *and all* $\bar{r}$ *the inequalities*

$$\mathcal{E}^\alpha(y, z, r, \bar{r}) \equiv h^\alpha(y, z, \bar{r}) - \bar{r}^\beta h_{r\beta}^\alpha(y, z, r) > 0 \qquad (\alpha = 1, \cdots, m)$$

hold unless $\bar{r} = kr,\ k \geqq 0$.

Our corollary is as follows:

THEOREM 13.1. *Let the functions* $h^\alpha(y, z, r)$ *be defined, continuous and positively homogeneous of degree* 1 *in* r *for all* (x, y, r), *and of class* C′ *for* $|r| > 0$. *Let hypotheses* (3.1), (3.2), (13.2), (13.3), (13.4) *and* (13.5) *be satisfied. Then in the class K there is a curve* C_0: $y = y(t), z = z(t)$ $(a \leqq t \leqq b)$ *which minimizes* $g(y(a), z(a), y(b), z(b))$.

The hypotheses of Theorem 13.1 guarantee the existence of a minimizing generalized curve

$$C_0^* : [y_0(t), z_0(t), \mathfrak{M}_0[t; \Phi(r, \rho)], M_0].$$

We suppose that t is the standard parameter. The function $F(y, z, r, \rho, \lambda)$ has the form

$$F = \lambda^\alpha(t, r, \rho)\{h^\alpha(y, z, r) - \rho^\alpha\}.$$

If as usual we denote by $S(t)$ the supporting set containing the vectors carried at $(y_0(t), z_0(t))$, by Lemma 13.1 of NC we find that each F_{ρ^i} is constant over $S(t)$. That is, by definition of F the multipliers λ^α for $S(t)$ are independent of r and ρ. By Lemma 13.2 of NC, these coincide with the multipliers of Theorem 10.1 of NC for all vectors carried. Hence for almost all t the λ^α are functions of t alone.

The DuBois-Reymond relations include the equations

$$F_{\rho^\alpha} = -\lambda^\alpha(t, r, \rho) = \psi^\alpha(t), \tag{13.6}$$

where we change the notation of GC slightly by writing

$$\begin{aligned} \psi^\alpha(t) &= c_\alpha + \int_0^t \mathfrak{M}_0[t; F_{z^\alpha}]\,dt \\ &= c_\alpha + \int_0^t \mathfrak{M}_0[t; \lambda^\beta h_{z^\alpha}^\beta]\,dt. \end{aligned} \tag{13.7}$$

Equations (13.6) hold (by Theorem 10.1 of NC) for all t in a set M comprising almost all of [0, 1] and for all r carried at $(y_0(t), z_0(t))$. Hence we may replace λ^α by $-\psi^\alpha(t)$ for all t, r and ρ without affecting any of the conclusions of Theorem 10.1 of NC. In particular, equation (13.7) takes the form

$$\psi^\alpha(t) = c_\alpha - \int_0^t \psi^\beta(t)\mu_\alpha^\beta(t)\,dt, \tag{13.8}$$

where for brevity we have written

$$\mu_\alpha^\beta(t) \equiv \mathfrak{M}_0[t; h_{z^\alpha}^\beta(y_0(t), z_0(t), r)]. \tag{13.9}$$

(This is defined on M_0; elsewhere we can define it arbitrarily, setting it equal say to 1.)

Equations (13.8) are equivalent to a set of differential equations, obtained by differentiating both members, together with the end conditions $\psi^\alpha(0) = c_\alpha$. Hence by a known property of such equations we can state

(13.10) *The solutions $\psi^\alpha(t)$ of equations* (13.8) *either never vanish simultaneously on the interval* $0 \leqq t \leqq 1$ *or else they are all identically zero.*

The curve C_0^* minimizes g on the subclass of K^* consisting of those curves

$$C^*\colon\ [y(t), z(t), \mathfrak{M}[t; \Phi(r, \rho)], M]$$

which satisfy the end conditions

$$\begin{aligned} y(a) - y_0(0) &= 0, \\ z(a) - z_0(0) &= 0, \\ y(b) - y_0(1) &= 0. \end{aligned} \tag{13.11}$$

Hence from the transversality conditions in Theorem 10.1 of NC we obtain

$$\lambda^0 g_{z_2^\alpha}(y_0(0), z_0(0), y_0(1), z_0(1)) + \psi^\alpha(1) = 0 \qquad (\alpha = 1, \cdots, m). \tag{13.12}$$

Since λ^0 is non-negative by Theorem 10.1 of NC, and the expression which it multiplies in (13.12) is non-negative by hypothesis (13.3), this implies

$$\psi^\alpha(1) \leqq 0 \qquad (\alpha = 1, \cdots, m). \tag{13.13}$$

Moreover, λ^0 must be positive. For if λ^0 is zero, all the $\psi^\alpha(1)$ vanish by (13.12), so all the $\psi^\alpha(t)$ vanish identically by (13.10). But now by (13.6) λ^0 and all the $\lambda^\alpha(t, r, \rho)$ vanish for all t in M and all (r, ρ) carried at $(y_0(t), z_0(t))$, and this contradicts Theorem 10.1 of NC. Since λ^0 is positive, there is no loss of generality in assuming that $\lambda^0 = 1$. Again by hypothesis (13.3), not all the derivatives $\partial g/\partial z_2^\alpha$ are zero, so by (13.12) not all the numbers $\psi^\alpha(1)$ are zero. We have thus shown

(13.14) *At least one of the numbers $\psi^\alpha(1)$ is negative.*

Next we establish a lemma concerning the solutions of equations (13.8).

LEMMA 13.1. *If the functions $\psi^\alpha(t)$ satisfy equations* (13.8), *and all the numbers $\psi^\alpha(1)$ are negative, and the functions $\mu^\alpha_\beta(t)$ $(\alpha \neq \beta;\ \alpha, \beta = 1, \cdots, m)$ are non-negative, then the functions $\psi^\alpha(t)$ are all negative on the entire interval* $0 \leqq t \leqq 1$.

Suppose that the lemma is false. There is then a number t_0 in the interval $[0, 1]$ such that some one of the $\psi^\alpha(t)$ vanishes at t_0, while all $\psi^\alpha(t)$ are negative for $t_0 < t \leqq 1$. We abandon the summation convention. For almost all t in $(t_0, 1)$ we have

$$\dot{\psi}^\alpha(t) = -\sum_{\beta=1}^{m} \psi^\beta(t)\mu^\beta_\alpha(t) \geqq -\psi^\alpha(t)\mu^\alpha_\alpha(t) \qquad (\alpha = 1, \cdots, m),$$

by (13.8) and the hypotheses. For such t the function $\psi^\alpha(t)$ is negative, so this implies

$$\dot{\psi}^\alpha(t)/\psi^\alpha(t) \leqq -\mu^\alpha_\alpha(t) \tag{13.15}$$

for almost all t such that $t_0 < t \leqq 1$. Choose a number τ such that $t_0 < \tau \leqq 1$, and integrate both members of inequality (13.15) from τ to 1. We obtain

$$\log|\psi^\alpha(1)| - \log|\psi^\alpha(\tau)| \leqq -\int_\tau^1 \mu^\alpha_\alpha\, dt \leqq \int_0^1 |\mu^\alpha_\alpha|\, dt. \tag{13.16}$$

This implies that $\log|\psi^\alpha(\tau)|$ has a finite lower bound, so that all the $\psi^\alpha(t)$ are bounded from zero on the interval $t_0 < t \leqq 1$. But this contradicts the assumption that one of these functions vanishes at t_0, and so the lemma is established.

As a corollary to Lemma 13.1 we have

LEMMA 13.2. *If the functions $\psi^\alpha(t)$ satisfy equations* (13.8), *and the numbers $\psi^\alpha(1)$ are non-positive and the functions $\mu^\alpha_\beta(t)$ $(\alpha \neq \beta;\ \alpha, \beta = 1, \cdots, m)$ are non-negative, then the functions $\psi^\alpha(t)$ are non-positive on the interval* $0 \leqq t \leqq 1$.

Let $\psi^\alpha(t, \epsilon)$ be the solutions of (13.8) with initial values $\psi^\alpha(1, \epsilon) = \psi^\alpha(1) - \epsilon$. By Lemma 13.1, the functions $\psi^\alpha(t, \epsilon)$ are negative on $[0, 1]$ for all positive ϵ. Since the solutions of the differential equations are continuous functions of the initial values, as ϵ tends to zero the negative-valued functions $\psi^\alpha(t, \epsilon)$ tend uniformly to $\psi^\alpha(t, 0) \equiv \psi^\alpha(t)$, so that these are non-positive.

We have already seen that if the hypotheses of Theorem 13.1 are satisfied, so are the hypotheses of Lemma 13.2. So by Lemma 13.2 and (13.10) we have

(13.17) *The functions $\psi^\alpha(t)$ are non-positive, and at no point t in* $[0, 1]$ *do they all vanish.*

The final step in our proof is to establish the following statement.

LEMMA 13.3. *For all t in M, there is only one vector (r_0, ρ_0) carried by C_0^* at $(y_0(t), z_0(t))$.*

Let (r_0, ρ_0) and (r, ρ) be non-null vectors carried at $(y_0(t), z_0(t))$. By (13.1) the vector r_0 is non-null; otherwise (r_0, ρ_0) would be null. Likewise r is non-null. By Theorem 10.1 of NC,

$$F_{r^i}(y_0(t), z_0(t), r_0, \rho_0, \lambda) = F_{r^i}(y_0(t), z_0(t), r, \rho, \lambda) \qquad (i = 1, \cdots, n).$$

Recalling that the multipliers $(\lambda^0, \lambda^\alpha)$ are $(1, -\psi^\alpha(t))$, we see that this becomes

$$-\psi^\alpha(t)h^\alpha_{r^i}(y_0(t), z_0(t), r_0) = -\psi^\alpha(t)h^\alpha_{r^i}(y_0(t), z_0(t), r).$$

If we multiply both members of this equation by r^i and sum, by virtue of the homogeneity of h^α we obtain

$$(13.18) \qquad -\psi^\alpha(t)\{h^\alpha(y_0(t), z_0(t), r) - r^i h^\alpha_{r^i}(y_0(t), z_0(t), r_0)\} = 0.$$

By Lemma 13.2 and hypothesis (13.5) each term in the sum (13.18) is non-negative, so each is zero. Not all the $\psi^\alpha(t)$ are zero; say $\psi^a(t) \neq 0$. Then the vanishing of the corresponding term in (13.18) implies

$$(13.19) \qquad h^a(y_0(t), z_0(t), r) - r^i h^a_{r^i}(y_0(t), z_0(t), r_0) = 0.$$

By hypothesis (13.5), there is a non-negative k such that $r = kr_0$. By equations (13.1) and the positive homogeneity of the h^α, it follows that $\rho = k\rho_0$. Hence (r, ρ) lies on the half-line $(kr_0, k\rho_0)$, $k \geqq 0$. Since (r_0, ρ_0) and (r, ρ) both have length $\mathfrak{L}(C_0^*)$, they are identical, and Lemma 13.3 is proved.

Now by definition C_0^* is an ordinary curve, so the track $y = y_0(t)$, $z = z_0(t)$ $(0 \leqq t \leqq 1)$ is a minimizing curve for g in the class K. The proof of Theorem 13.1 is complete.

14. **A second theorem on Mayer problems.** Another theorem similar in type to Theorem 13.1 is the following.

THEOREM 14.1. *Let the functions $h^\alpha(y, z, r)$ be defined, continuous and positively homogeneous of degree* 1 *in r, and of class* C′ *for $|r| > 0$. Let hypotheses* (1.3) *and* (13.2) *be satisfied, and also the following three hypotheses.*

(14.1) *For each point in P all the partial derivatives $\partial g/\partial z_2^\alpha$ are defined, continuous and positive.*

(14.2) *For all (y, z, r) the partial derivatives*

$$h^\alpha_{z^\beta}(y, z, r) \qquad (\alpha \neq \beta;\ \alpha, \beta = 1, \cdots, m)$$

are non-negative.

(14.3) *For all (y, z, r) with $|r| > 0$ and all $\bar r$ the inequality*

$$\mathcal{E}^\alpha(y, z, r, \bar r) \equiv h^\alpha(y, z, \bar r) - \bar r^\beta h^\alpha_{r^\beta}(y, z, r) \geqq 0$$

is satisfied.

Then in the class K there is a curve C which minimizes $g(y(a), z(a), y(b), z(b))$.

As before, equations (13.12) hold, and imply $\lambda^0 > 0$. But now (13.12) and (14.1) imply that

$$(14.4) \qquad \psi^\alpha(1) < 0 \qquad (\alpha = 1, \cdots, m).$$

From hypotheses (14.2) we deduce as before that

$$\mu^\alpha_\beta(t) \geqq 0 \qquad (\alpha \neq \beta;\ \alpha, \beta = 1, \cdots, m;\ 0 \leqq t \leqq 1).$$

Now by Lemma 13.1 all the functions $\psi^\alpha(t)$ are negative for all t in [0, 1].

As a substitute for Lemma 13.3 we prove

LEMMA 14.1. *For all t in M, the supporting set $S(t)$ is its own first (and last) portion.*

Let (r_0, ρ_0) and (r, ρ) be non-null vectors in $S(t)$. As before, equation (13.18) holds, and each term in (13.18) vanishes. But no factor $\psi^\alpha(t)$ is zero, so this implies

$$(14.5) \qquad h^\alpha(y_0(t), z_0(t), r) - r^i h^\alpha_{r^i}(y_0(t), z_0(t), r_0) = 0 \qquad (\alpha = 1, \cdots, m).$$

By hypothesis (14.3), the function $h^\alpha(y_0(t), z_0(t), r)$ is a convex function of r. By (14.5), if (r, ρ) is in $S(t)$, the vector r lies in the set K_α (necessarily convex) on which the convex function h^α coincides with the supporting linear function

$$r^i h^\alpha_{r^i}(y_0(t), z_0(t), r_0).$$

Hence all solutions r of (14.5) lie in the convex product set

$$K_0 \equiv K_1 K_2 \cdots K_m .$$

We now define K to be the set of all vectors (r, ρ) such that r is in K_0 and equations (13.1) are satisfied. Since the h^α are linear on K_0, the set K is the linear image of the convex set K_0, and is therefore convex. On it the equations (13.1) hold, by its definition, and on it the integrand f is convex, being identically zero. We have seen that if (r, ρ) is in $S(t)$ equations (14.5) hold, so that r is in K_0 and (r, ρ) in K. Hence $S(t)$ is contained in K. Finally, $S(t)$ is its own first and last portion, and the lemma is established.

Now the hypotheses of Theorem 6.1 have all been shown satisfied, so the conclusion of that theorem holds. This completes the proof of Theorem 14.1.

15. **Continuity properties of the solutions of Mayer problems.** The hypotheses of Theorem 13.1 imply more than the existence of a solution; they also guarantee certain continuity properties of the solution. For the sake of a slight gain in generality we state these properties in a separate theorem.

THEOREM 15.1. *Let hypotheses* (1.1), (3.1), (13.2), (13.3), (13.4) *and* (13.5) *be satisfied. Let $C_0 : y = y_0(t), z = z_0(t)$ $(0 \leqq t \leqq 1)$ be the standard representation of a curve interior to E which furnishes a strong relative minimum for $g(y(a), z(a), y(b), z(b))$ in the class K. Then the functions $y(t)$ and $z(t)$ and the multipliers $\lambda^\alpha(t)$ have continuous first derivatives, and the curve C_0 is normal. Moreover, if the functions h^α are of class C^κ $(\kappa > 1)$ for $|r| > 0$ and the inequalities*

$$h^\alpha_{r^i r^j}(y_0(t), z_0(t), y_0'(t))\xi^i \xi^j > 0 \qquad (\alpha = 1, \cdots, m)$$

all hold except when the vector ξ is a multiple $ky_0'(t)$ of y_0', the functions $y_0(t)$, $z_0(t)$ and $\lambda^\alpha(t)$ are of class C^κ.

By Theorem 16.1 of NC, there are bounded measurable functions $\lambda^\alpha(t)$ such that the function

$$F(y, z, r, \rho, \lambda) \equiv \lambda^\alpha(t)\{h^\alpha(y, z, r) - \rho^\alpha\}$$

satisfies the conclusions of that theorem. In particular, for all t in a subset M of [0, 1] with measure 1 the DuBois-Reymond relations hold, whence

$$F_{\rho^\alpha} = -\lambda^\alpha(t) = \psi^\alpha(t) \equiv c_\alpha + \int_0^t \lambda^\beta(t) h^\beta_{z^\alpha}(y_0, z_0, \dot{y}_0)\, dt. \tag{15.1}$$

This permits us to assume

$$\lambda^\alpha(t) = -\psi^\alpha(t) \qquad (0 \leqq t \leqq 1, \alpha = 1, \cdots, m); \tag{15.2}$$

this may alter the definition of λ^α on a set of measure 0, but such a change is of no importance. We define

$$\mu^\alpha_\beta(t) = h^\alpha_{z^\beta}(y_0(t), z_0(t), \dot{y}_0(t)). \tag{15.3}$$

As in §13, we find that the $\psi^\alpha(t)$ are non-positive and never vanish simultaneously.

In (15.1) we used m of the $\nu = m + n$ DuBois-Reymond equations. The others are

$$F_{r^i} = \lambda^\alpha(t) h^\alpha_{r^i}(y_0(t), z_0(t), \dot{y}_0(t)) = c_i + \int_0^t \lambda^\alpha h^\alpha_{y^i}\, dt, \tag{15.4}$$

valid for all t in M. We now show that the function $y_0'(t)$, regarded as defined on M, is uniformly continuous. Suppose this false. It is then possible to find two sequences t_n, $\bar{t}_n$ of points of M converging to a limit point t_0 in [0, 1] (not necessarily in M) such that

$$\lim_{n\to\infty} y_0'(t_n) = r, \qquad \lim_{n\to\infty} y_0'(\bar{t}_n) = \bar{r} \neq r. \tag{15.5}$$

On M the vector y_0' has length $\mathfrak{L}(C_0)$, so from (15.5) we at once deduce

$$| r | = | \bar{r} | = \mathfrak{L}(C_0). \tag{15.6}$$

The functions λ^α are continuous by (15.1) and (15.2); hence by (15.4) and (15.5)

$$\begin{aligned} \lambda^\alpha(t_0) h^\alpha_{r^i}(y_0(t_0), z_0(t_0), r) &= \lim_{n\to\infty} \lambda^\alpha(t_n) h^\alpha_{r^i}(y_0(t_n), z_0(t_n), y_0'(t_n)) \\ &= c_i + \int_0^{t_0} \lambda^\alpha h^\alpha_{z^i}\, dt; \end{aligned} \tag{15.7}$$

and likewise

$$\lambda^\alpha(t_0) h^\alpha_{r^i}(y_0(t_0), z_0(t_0), \bar{r}) = c_i + \int_0^{t_0} \lambda^\alpha h^\alpha_{z^i}\, dt. \tag{15.8}$$

Thus the left members of (15.7) and (15.8) are equal. If we subtract them and multiply by $\bar{r}_i$, we find

$$\lambda^\alpha(t_0)\{h^\alpha(y_0(t_0), z_0(t_0), \bar{r}) - \bar{r}^i h^\alpha_{r^i}(y_0(t_0), z_0(t_0), r)\} = 0. \tag{15.9}$$

The ψ^α are non-positive, so the λ^α are non-negative. The factors in braces in

(15.9) are non-negative, by (14.5). Hence each term must vanish. Some one $\lambda^\alpha(t_0)$, say $\lambda^a(t_0)$, is not zero; hence

$$h^a(y_0(t_0), z_0(t_0), \bar{r}) - \bar{r}^i h^a_{r^i}(y_0(t_0), z_0(t_0), r) = 0. \tag{15.10}$$

By (13.5), this implies that $\bar{r} = Kr$, $k \geqq 0$. Recalling (15.6), we see that $\bar{r} = r$, and this contradicts the original assumptions that $\bar{r} \neq r$. We have therefore shown that y_0' is uniformly continuous on M.

From this it follows that there exists a function $r(t)$ continuous on [0, 1] and equal to $y_0'(t)$ on M. Hence

$$y^i(t) = y^i(0) + \int_0^t r^i(t)\, dt,$$

so that $y^{i\prime}(t)$ is everywhere equal to the continuous function $r(t)$.

From the differential equations (13.1), written in the form

$$z_0^{\alpha\prime}(t) = h^\alpha(y_0(t), z_0(t), y_0'(t)), \tag{15.11}$$

we see that z_0' also is continuous. From the differential equations (15.1) it follows that $\psi^\alpha(t)$ also has a continuous first derivative. In fact, if we anticipate a part of the proof of the last sentence of our theorem, if the functions y_0 and z_0 are of class $C^{(\kappa)}$, the functions $\mu^\alpha_\beta(t)$ are of class $C^{(\kappa-1)}$, because of (15.3); so by known properties of solutions of linear differential equations the functions $\psi^\alpha(t)$ will be of class $C^{(\kappa)}$, and so will their negatives, the $\lambda^\alpha(t)$.

In the proof in §13 that λ^0 is not zero we used only the DuBois-Reymond and transversality conditions. We therefore have shown that there are no multipliers $(0, \lambda^\alpha(t))$ with which C_0 satisfies those conditions. That is, C_0 is a normal curve.

Let us temporarily return to the notation of §§1–8. The matrix

$$R \equiv \begin{Vmatrix} F_{r^i r^j} & \varphi^\alpha_{r^i} \\ \varphi^\beta_{r^j} & 0 \end{Vmatrix}, \tag{15.12}$$

in which the arguments are $(y_0(t), y_0'(t))$, is necessarily singular; if we multiply its $\nu + m$ columns by coefficients $(c_1, \cdots, c_{\nu+m})$ proportional to $(y^{1\prime}, \cdots, y^{\nu\prime}, 0, \cdots, 0)$ the combination is a null vector. If no other linear combination of the column vanishes, R has rank $\nu + m - 1$. In this case the curve C_0 is said to be non-singular. From the classical theory of the calculus of variations it is well known that if C_0 is of class C' and is non-singular, and the functions $f(y, r)$ and $\varphi^\alpha(y, r)$ are of class $C^{(\kappa)}$, then C_0 must be of class $C^{(\kappa)}$. To complete the proof of our theorem it is then sufficient to show that when the hypotheses in the last sentence of Theorem 15.1 are satisfied, the curve C_0 is non-singular.

For the present problem, the ρ^α enter F linearly, so the partial derivatives $F_{r^i \rho^\alpha}$ and $F_{\rho^\alpha \rho^\beta}$ are all zero. The matrix R of (15.12) takes the form

$$R = \begin{Vmatrix} 0 & 0 & -\delta_{\beta\alpha} \\ 0 & F_{r^i r^j} & h^\alpha_{r^i} \\ -\delta_{\gamma s} & h^\gamma_{r^j} & 0 \end{Vmatrix}, \tag{15.13}$$

where $\delta_{\beta\alpha}$ is the Kronecker delta. Let $C_1, \cdots, C_\nu, d_1, \cdots, d_n, e_1, \cdots, e_\nu$ be coefficients such that on multiplying each of the columns of R by the corresponding coefficient and adding the sum is zero. From the vanishing of the upper ν components of the sum we see that the e_α all are zero. The vanishing of the next m components implies

$$F_{r^i r^j} d_j = 0 \qquad (i = 1, \cdots, n).$$

We multiply by d_i and add. On recalling the definition of F, we find that

$$\lambda^\alpha(t) h^\alpha_{r^i r^j}(y_0, z_0, y_0') d_i d_j = 0. \tag{15.14}$$

The λ^α are non-negative; and so are the quadratic forms they multiply, by hypothesis. Hence each term vanishes. At least one of the λ^α, say $\lambda^a(t)$, is not zero; hence

$$h^a_{r^i r^j}(y_0, z_0, y_0') d_i d_j = 0. \tag{15.15}$$

By hypothesis, this implies that there is a k such that

$$d_j = k y_0^{j\prime}(t) \qquad (j = 1, \cdots, n). \tag{15.16}$$

Turning to the last ν components of the linear combination, using the homogeneity of the h^α, and recalling (15.11), we find

$$\begin{aligned} 0 &= C_\epsilon(-\delta_{\gamma\epsilon}) + d_j h^\gamma_{r^j} \\ &= -C_\gamma + k y_0^{j\prime}(t) h^\gamma_{r^j}(y_0, z_0, y_0') \\ &= -C_\gamma + k h^\gamma(y_0, z_0, y_0') \\ &= -C_\gamma + k z_0^{\gamma\prime}(t). \end{aligned}$$

Therefore

$$C_\gamma = k z_0^{\gamma\prime}(t). \tag{15.17}$$

This, with (15.16), shows that the multipliers $(C_\gamma, d_j, e_\alpha)$ are necessarily proportional to $(z_0', y_0', 0)$, and the matrix R of (15.13) has rank $\nu + m - 1$. We have thus shown that C_0 is non-singular and completed the proof of Theorem 15.1.

16. **A remark concerning the determination of supporting sets.** In order to verify the hypotheses of Theorems 6.1, 7.1 and 8.1 we must find the sets S which are supporting sets at given points y. The finding of these sets is made more difficult by the dependence of the multipliers $\lambda^\alpha(r)$ of the set S (which in Theorems 6.1 and 7.1 are the multipliers $\lambda^\alpha(t, r)$ for C_0^*) on the variables r. Accordingly, we shall show that under certain conditions these multipliers are actually independent of r.

If at a point y_0 one component r^k of r enters linearly, with a non-zero coefficient, in one of the functions $\varphi^\alpha(y, r)$ (say $\varphi^a(y, r)$), the equation $\varphi^a(y_0, r) = 0$ can be

solved for r^k and used to eliminate r^k from $f(y_0, r)$ and all the functions $\varphi^\alpha(y_0, r)$, $\alpha \neq a$. We thus have the situation:

(16.1) *At $y = y_0$ the function $\varphi^a(y, r)$ has the form $hr^k + g(r)$, where $h \neq 0$ and $g(r)$ is independent of r^k.*

(16.2) *At $y = y_0$ all the functions $f(y, r)$ and $\varphi^\alpha(y, r)$ $(\alpha = 1, \cdots, a - 1, a + 1, \cdots, m)$ are independent of r^k.*

Our aid to computation is

THEOREM 16.1. *If the functions $f(y, r)$, $\varphi^\alpha(y, r)$ satisfy conditions* (16.1) *and* (16.2), *and S is a supporting set at y_0, with multipliers $\lambda^\alpha(r)$, the multiplier $\lambda^a(r)$ is independent of r.*

As usual, we define

$$F(y, r, \lambda) = \lambda^0 f(y, r) + \lambda^\alpha(r)\varphi^\alpha(y, r). \tag{16.3}$$

By the definition of supporting set, there is a constant c such that the equation

$$F_{r^k}(y_0, r, \lambda) = c \tag{16.4}$$

holds for all non-null vectors in S. But by (16.1) and (16.2), equation (16.4) has the form

$$\lambda^a(r)h = c. \tag{16.5}$$

Since $h \neq 0$, this implies that $\lambda^a(r)$ has the same value c/h for all non-null vectors in S.

Another remark on the application of our theorems concerns finite side conditions. If one of the $\varphi^\alpha(y, r)$, say $\varphi^a(y, r)$, is independent of the r, it can be replaced by the equations

$$\varphi^a_{y^i}(y)r^i = 0, \tag{16.6}$$

the initial condition $\varphi^a(y(a)) = 0$ being imposed. But this device causes trouble in our theorems. For if we choose $\lambda^a = 1$ and choose all other λ's equal to 0, we find that the whole r-space is a supporting set and Ω vanishes identically. Hence if (16.6) is included among the side conditions, no point y can be ordinary unless it happens that the solutions of the other equations

$$\varphi^\alpha(y_0, r) = 0 \qquad (\alpha = 1, \cdots, a - 1, a + 1, \cdots, m)$$

form a convex set on which $f(y_0, r)$ is convex.

As a result, if the side conditions include equations independent of the r, it seems desirable to use these equations to eliminate some of the y and their derivatives, thus transforming the problem into one with fewer variables and no finite side conditions. This will be illustrated in the next section.

17. **An example.** We conclude with an example in which there are three side conditions, one of which is a differential equation, one a finite equation and

one an isoperimetric condition. In this problem, as is usual in brachistochrone problems, the integrand has an infinite discontinuity, and also it is not quite trivially evident that condition (3.2) is satisfied. Briefly, we seek a curve joining two fixed points, lying in a surface immersed in a resisting medium, having an assigned moment $\int x\,ds$ and assigned initial and terminal velocities, for which the time of descent of a particle is the shortest possible.

Let Σ be a surface defined by an equation

$$y = Y(x, z), \tag{17.1}$$

where $Y(x, z)$ is defined and of class C'' for all x and z.

The effect of the resisting medium is to produce a frictional retardation $R(v)$ depending on the velocity v. We assume that

(17.2) *$R(v)$ is defined and continuous for $v \geqq 0$ and has a continuous derivative $R'(v)$ for all $v > 0$;*

also, we assume that

(17.3) *$R(v) > 0$ and $R'(v) \geqq 0$ for all $v > 0$.*

If we choose our units so that the acceleration due to gravity is 1, a particle traversing the curve

$$C: x = x(t),\ y = y(t),\ z = z(t) \qquad (a \leqq t \leqq b) \tag{17.4}$$

under gravity has a velocity $v(t)$ which satisfies the equation[17]

$$v\dot{v} + \dot{z} + R(v)[\dot{x}^2 + \dot{y}^2 + \dot{z}^2]^{\frac{1}{2}} = 0, \tag{17.5}$$

the z-axis being supposed vertical. The time of descent of the particle is given by

$$\mathcal{T}(C) = \int_a^b v^{-1}[\dot{x}^2 + \dot{y}^2 + \dot{z}^2]^{\frac{1}{2}}\,dt. \tag{17.6}$$

Consider now the problem of determining that curve which joins two fixed points of the surface Σ for which the time of descent of a particle is least, subject to the following conditions.

(17.7) *The curve lies in the surface Σ.*

(17.8) *The initial and final velocities of the particle have assigned values.*

(17.9) *Equation* (17.5) *is satisfied.*

(17.10) *The integral*

$$\int_a^b x[\dot{x}^2 + \dot{y}^2 + \dot{z}^2]^{\frac{1}{2}}\,dt$$

has an assigned value γ.

[17] Bolza, *Vorlesungen über Variationsrechnung*, p. 577.

For this problem we prove that *if the class K of curves satisfying conditions* (17.7), (17.8), (17.9) *and* (17.10) *contains a curve for which $\mathcal{F}(C)$ is finite, it contains a curve which minimizes $\mathcal{F}(C)$.*

Our first step is to transform the problem so as to rid ourselves of the finite condition (17.7). From (17.1) we obtain

$$\dot{y} = Y_x\dot{x} + Y_z\dot{z}, \tag{17.11}$$

so that

$$[\dot{x}^2 + \dot{y}^2 + \dot{z}^2]^{\frac{1}{2}} = \varphi(x, z, \dot{x}, \dot{z}) \equiv [\dot{x}^2 + \dot{z}^2 + (Y_x\dot{x} + Y_z\dot{z})^2]^{\frac{1}{2}}. \tag{17.12}$$

The function φ is positive unless $\dot{x}$ and $\dot{z}$ vanish, and its $\mathcal{E}$-function is positive except for the trivial vanishing. Our problem is now to minimize

$$\mathcal{F}(C) \equiv \int_a^b v^{-1}\varphi(x, z, \dot{x}, \dot{z})\, dt$$

subject to fixed end values for x, y, and v, to the isoperimetric condition

$$\int_a^b x\varphi\, dt = \gamma, \tag{17.13}$$

and to the differential equation

$$v\dot{v} + \dot{z} + R(v)\varphi(x, z, \dot{x}, \dot{z}) = 0. \tag{17.14}$$

This is a problem in three-space, with coördinates (x, z, v). We replace $(\dot{x}, \dot{z}, \dot{v})$ notationally by (ξ, ζ, η).

The corresponding generalized-curve problem is to minimize

$$\mathcal{F}(C^*) = \int_a^b \mathfrak{M}[t; v^{-1}\varphi(x(t), z(t), \xi, \zeta)]\, dt$$

in the class K^* of generalized curves

$$C^*\colon\ [x(t), z(t), v(t), \mathfrak{M}[t; \Phi(\xi, \zeta, \eta)], M]$$

which satisfy the following conditions.

(17.15) *The ends $(x(a), z(a), v(a))$ and $(x(b), z(b), v(b))$ have the values required for curves of K.*

(17.16) *For almost all t, the equation*

$$v\eta + \zeta + R(v)\varphi(x(t), z(t), \xi, \zeta) = 0$$

holds for all vectors carried at t.

$$\int_a^b \mathfrak{M}[t; x\varphi]\, dt = \gamma. \tag{17.17}$$

We shall show that condition (3.2) is satisfied. This proof uses only equation (17.16) and the boundedness of $x(a)$, $y(a)$, $z(a)$ and $v(a)$; omission of the require-

ment that the track of C^* lie in the surface (Σ) would merely simplify the proof slightly.

On integration, equation (17.16) yields (with (17.3))

$$\int^t \mathfrak{M}[t; v\eta]\,dt + \int_a^t \mathfrak{M}[t; \zeta]\,dt \leqq 0.$$

When we recall that $\mathfrak{M}[t; v\eta] = vv'$ and $\mathfrak{M}[t; \zeta] = z'$ for almost all t, this yields

$$\tfrac{1}{2}\{v^2(t) - v^2(a)\} \leqq -[z(t) - z(a)]. \tag{17.18}$$

If we define

$$l(t) = \int_a^t \mathfrak{M}[t; \varphi(x(t), z(t), \xi, \zeta)]\,dt, \tag{17.19}$$

we readily prove

$$l(t) \geqq | z(t) - z(a) |,$$

so that (17.18) implies

$$v(t) \leqq [v^2(a) + 2l(t)]^{\frac{1}{2}}. \tag{17.20}$$

Substituting this in the definition of $\mathcal{F}(C^*)$ gives

$$\begin{aligned} \mathcal{F}(C^*) &= \int_a^b \mathfrak{M}[t; v^{-1}\varphi(x(t), z(t), \xi, \zeta)]\,dt \\ &= \int_a^b v^{-1} l(t)\,dt \\ &\geqq \int_a^b [v^2(a) + 2l(t)]^{-\frac{1}{2}} l(t)\,dt \\ &= [v^2(a) + 2l(b)]^{\frac{1}{2}} - v(a). \end{aligned} \tag{17.21}$$

From (17.20) and (17.21) we deduce

(17.22) *On every subclass of K^* on which $v(a)$ and $\mathcal{F}(C^*)$ are bounded, $l(t)$ and $v(t)$ are also bounded.*

From the definition of φ we at once see that

$$\xi^2 + \zeta^2 \leqq [\varphi(x, z, \xi, \zeta)]^2. \tag{17.23}$$

This, with (17.16), yields

$$| \eta | = v^{-1} | \zeta + R(v)\varphi(x, z, \xi, \zeta) | \leqq v^{-1}(R(v) + 1)\varphi. \tag{17.24}$$

Combining this with (17.23) gives us

$$\xi^2 + \zeta^2 + \eta^2 \leqq \{1 + v^{-2}(R(v) + 1)^2\}\varphi^2, \tag{17.25}$$

or, since $a^{\frac{1}{2}} + b^{\frac{1}{2}} \geqq (a + b)^{\frac{1}{2}}$,

$$[\xi^2 + \zeta^2 + \eta^2]^{\frac{1}{2}} \leqq [1 + v^{-1}(R(v) + 1)]\varphi. \tag{17.26}$$

By (17.22), on every minimizing sequence $l(t)$ and $R(v(t))$ are bounded. Choose a minimizing sequence, and let M be an upper bound for $l(t)$ and $R(v(t)) + 1$ on all the curves of the sequence. From (17.26) we obtain

$$
\begin{aligned}
\mathfrak{L}(C^*) &= \int_a^b \mathfrak{M}[t; \{\xi^2 + \zeta^2 + \eta^2\}^{\frac{1}{2}}]\, dt \\
&\leqq \int_a^b \mathfrak{M}[t; \varphi]\, dt + \int_a^b (R(v) + 1)\mathfrak{M}[t; v^{-1}\varphi]\, dt \\
&\leqq M + M\mathcal{F}(C^*).
\end{aligned} \tag{17.27}
$$

Hence the curves of the minimizing sequence have uniformly bounded lengths, and condition (3.2) is satisfied.

It is now possible, as in §8 of GC, to choose a convergent minimizing sequence $\{C_n^*\}$ in K^*, with a limit curve C_0^* in K^*. Let C_0^* be given in standard representation by the formula

$$C_0^*: [x_0(t), z_0(t), v_0(t), \mathfrak{M}_0[t; \psi], M_0].$$

We cannot at once conclude that

$$\mathcal{F}(C_0^*) = \mu \equiv \lim_{n\to\infty} \mathcal{F}(C_n^*), \tag{17.28}$$

because of the discontinuity of the integrand of $\mathcal{F}(C^*)$ at $v = 0$. To avoid this difficulty we define

$$\mathcal{F}_m(C^*) = \int_a^b \mathfrak{M}[t; \{\max(v, 1/m)\}^{-1} \varphi(x, z, \xi, \zeta)]\, dt \qquad (m = 1, 2, 3, \cdots). \tag{17.29}$$

The integrand here is continuous, and does not exceed that of $\mathcal{F}$. Hence

$$\mathcal{F}_m(C_0^*) = \lim_{n\to\infty} \mathcal{F}_m(C_n^*) \leqq \mu \qquad (m = 1, 2, \cdots). \tag{17.30}$$

As m increases, the integrand of $\mathcal{F}_m$ is non-decreasing. So by a familiar theorem in Lebesgue integration theory,

$$\mathcal{F}(C_0^*) = \lim_{m\to\infty} \mathcal{F}_m(C_0^*) \leqq \mu.$$

The curve C_0^* is in K^*, so by its definition the number μ cannot be less than $\mathcal{F}(C_0^*)$, and equation (17.28) is satisfied. We have therefore established the existence of a minimizing generalized curve for $\mathcal{F}(C^*)$ in the class K^*.

Since $\varphi(x, z, \xi, \zeta)$ is positive for every vector carried, by (17.12), at each t of M_0 such that $v_0(t) = 0$ we have

$$\mathfrak{M}[t; \{\max(v_0(t), m^{-1})\}^{-1}\varphi(x_0, z_0, \xi, \zeta)] = m\mathfrak{M}[t; \varphi] \to \infty. \tag{17.31}$$

Hence inequality (17.30) implies that $v_0(t)$ is positive for almost all t in M_0. The set of t on which $v_0(t)$ vanishes is closed. Hence the part of C_0^* on which v_0 is positive can be regarded as the sum of a denumerable set of arcs of C_0^*, corresponding to intervals $\alpha_i \leqq t \leqq \beta_i$ ($i = 1, 2, \cdots$), along each of which the

function $v_0(t)$ is positive. Our task of showing that C_0^* is an ordinary curve will therefore be complete as soon as we have shown that each arc $\alpha \leqq t \leqq \beta$ of C_0^* along which v_0 is positive is an arc of an ordinary curve.

Consider then such an arc. If (ξ, ζ, η) satisfies (17.16) and is not (0, 0, 0), then (ξ, ζ) is not (0, 0), so near (ξ, ζ, η) the left member of (17.16) and the integrand in $\mathcal{F}(C^*)$ are of class C′. Thus hypothesis (1.1) holds. Condition (1.2) is satisfied, as is easily verified. If now we can show that at each t in $[\alpha, \beta]$ the supporting set $S(t)$ has a first or last portion, by Theorem 6.1 the arc $\alpha \leqq t \leqq \beta$ of C_0^* will be an ordinary curve, and our statement that a minimizing curve exists will be justified.

Consider therefore a point t in $[\alpha, \beta]$. The supporting set $S(t)$ has multipliers λ_0, λ_1, $\lambda_2(t, \xi, \zeta, \eta)$ such that the partial derivatives with respect to ξ, ζ and η of

$$\begin{aligned} F(x, z, v, \xi, \zeta, \eta, \lambda) \equiv \lambda_0 v^{-1} \varphi(x, z, \xi, \zeta) + \lambda_1 x\varphi \\ + \lambda_2(t, \xi, \zeta, \eta)[v\eta + \zeta + R(v)\varphi] \end{aligned} \tag{17.32}$$

are constant on the class of all non-null vectors of $S(t)$. By Theorem 16.1, $\lambda_2(t, \xi, \zeta, \eta)$ is independent of ξ, ζ and η; we redesignate it by $\lambda_2(t)$. There are two cases to be distinguished.

Case 1. *The equation*

$$\lambda_0 v^{-1} + \lambda_1 x + \lambda_2(t)R(v) = 0 \tag{17.33}$$

is not satisfied.

Let (ξ, ζ, η) and (ξ_0, ζ_0, η_0) be non-null vectors belonging to $S(t)$. Then the partial derivatives of F with respect to ξ, ζ and η have the same values for these two sets, so that

$$\varphi_\xi(x, z, \xi, \zeta) = \varphi_\xi(x, z, \xi_0, \zeta_0), \qquad \varphi_\zeta(x, z, \xi, \zeta) = \varphi_\zeta(x, z, \xi_0, \zeta_0).$$

Since φ is positively homogeneous of degree 1, this implies

$$\varphi(x, z, \xi, \zeta) - \xi\varphi_\xi(x, z, \xi_0, \zeta_0) - \zeta\varphi_\zeta(x, z, \xi_0, \zeta_0) = 0.$$

The left member of this equation is the $\mathcal{E}$-function for φ, and its vanishing implies that $\xi_0 = k\xi$ and $\zeta_0 = k\zeta$, $k > 0$. This, with (17.16), implies that $\eta_0 = k\eta$. It follows that the entire set $S(t)$ consists of a half-line $(k\xi, k\zeta, k\eta)$, $k \geqq 0$. It is then its own first and last portion.

Case 2. *The equation* (17.33) *is satisfied.*

If we compute the partial derivatives of F with respect to ξ, ζ and η, we see that they are constants. Therefore the set $S(t)$ contains every vector (ξ, ζ, η) such that $(x(t), z(t), v(t), \xi, \zeta, \eta)$ is admissible. We shall show that $S(t)$ contains a last portion.

Because of the homogeneity of the Ω-function, it is sufficient to compute it for vectors such that

$$\varphi(x, z, \xi, \zeta) = \varphi(x, z, \xi_0, \zeta_0) = 1. \tag{17.34}$$

Under this assumption, recalling (17.16) and (17.33) we compute

$$\begin{aligned}
&\Omega(x, z, v, \xi, \zeta, \eta, \xi_0, \zeta_0, \eta_0, \lambda) \\
(17.35)\qquad &= (-\lambda_0 v^{-2} + \lambda_2(t)R'(v))(\eta - \eta_0) + \lambda_1(\xi - \xi_0) \\
&= (\lambda_0 v^{-2} - \lambda_2(t)R'(v))v^{-1}(\zeta - \zeta_0) + \lambda_1(\xi - \xi_0).
\end{aligned}$$

It is not possible that the equations

$$(17.36)\qquad \lambda_1 = 0,$$

$$(17.37)\qquad \lambda_0 v^{-2} - \lambda_2(t)R'(v) = 0$$

both hold. For if (17.36) holds, equations (17.33) (with $\lambda_1 = 0$) and (17.37) are linear homogeneous equations for λ_0 and λ_2, and their determinant must vanish, since λ_0, λ_1, and $\lambda_2(t)$ cannot all vanish at any t. But this determinant is

$$v^{-1}R'(v) + v^{-2}R(v),$$

and the first term is non-negative and the second positive, so it cannot vanish.

Now, since the coefficients of $\zeta - \zeta_0$ and $\xi - \xi_0$ do not both vanish, as in §11 (from (11.8) on) we see that there is a vector (ξ, ζ) such that the final member of (17.35) is positive unless $(\xi_0, \zeta_0) = (\xi, \zeta)$. If we determine η by (17.16), the half-line $(k\xi, k\zeta, k\eta)$ through the vector thus determined is the last portion of $S(t)$.

Thus in both cases $S(t)$ has a last portion, and the proof of the existence of a minimizing curve is complete.

Of course a minimizing curve still exists if the conditions are relaxed, provided that we still demand the boundedness of $x(a)$, $y(a)$, $z(a)$ and $v(a)$. If we omit the isoperimetric condition (17.13) and leave $v(b)$ free, we have the well-known problem of the brachistochrone on a surface in a resisting medium. The requirement that the curve lie in the surface (17.1) may also be dispensed with. If we dispense with the resisting medium, the analysis fails, since (17.3) is not satisfied. But then v can be eliminated from the integrand in $\mathcal{F}$, and we get the familiar form

$$\mathcal{F}(C) = \int_a^b [(\dot{x}^2 + \dot{y}^2 + \dot{z}^2)/(\alpha - z)]^{\frac{1}{2}}\, dt.$$

To this the analysis in §11 applies, even if we impose, say, the isoperimetric conditions

$$\int_a^b x[\dot{x}^2 + \dot{y}^2 + \dot{z}^2]^{\frac{1}{2}}\, dt = \gamma,$$

$$\int_a^b [\dot{x}^2 + \dot{y}^2 + \dot{z}^2]^{\frac{1}{2}}\, dt = l.$$

THE UNIVERSITY OF VIRGINIA.

AN ANALOGUE OF THE STAUDT-CLAUSEN THEOREM

By L. Carlitz

1. A set of rational functions B_m of an indeterminate x is defined by means of

$$\frac{t}{\psi(t)} = \sum_{m=0}^{\infty} \frac{B_m}{g_m} t^m,$$

where the several quantities involved are defined as follows. Put

$$[k] = x^{p^{nk}} - x,$$
$$F_k = [k][k-1]^{p^n} \cdots [1]^{p^{n(k-1)}}, \qquad F_0 = 1,$$
$$L_k = [k][k-1] \cdots [1], \qquad L_0 = 1;$$

the function in the denominator is given by[1]

$$\psi(t) = \sum_{k=0}^{\infty} \frac{(-1)^k}{F_k} t^{p^{nk}}.$$

Further for

$$m = \alpha_0 + \alpha_1 p^n + \cdots + \alpha_s p^{ns} \qquad (0 \leqq \alpha_i < p^n),$$

write

$$g_m = F_0^{\alpha_0} F_1^{\alpha_1} \cdots F_s^{\alpha_s}, \qquad g_0 = 1.$$

Thus B_m is defined for all $m \geqq 0$ and vanishes if m is not a multiple of $p^n - 1$. From the formula

$$\sum \frac{1}{E^m} = \frac{B_m}{g_m} \xi^m \qquad (p^n - 1 \mid m),$$

where the summation extends over all primary polynomials $E = E(x)$ with coefficients in $GF(p^n)$, and where

$$\xi = \lim_{k=\infty} \frac{(x^{p^n} - x)^{p^{nk}/(p^n-1)}}{L_k},$$

it follows that

$$B_m \neq 0 \qquad \text{for } p^n - 1 \mid m.$$

In discussing properties of B_m we therefore assume that m is a multiple of $p^n - 1$. Note that the coefficients involved in B_m lie in $GF(p)$, that is, are integers (mod p).

Received October 10, 1940.

[1] See this Journal, vol. 1(1935), pp. 137–168.

The following theorem on the fractional part of B_m—analogous to the Staudt-Clausen theorem for the Bernoulli numbers—was proved some time ago:[2]

For given $m > 0$, the system

$$(1.1)\qquad \begin{aligned} m &= \sum_{i=0}^{s-1} \alpha_i p^{nki}, & p^{nk} - 1 &= \sum_{i=0}^{s-1} \alpha_i, & \\ \alpha_i &= \sum_{j=0}^{nk-1} \alpha_{ij} p^j, & p - 1 &= \sum_{i=0}^{s-1} \alpha_{ij}, & \alpha_{ij} &\geqq 0, \end{aligned}$$

is either (i) *inconsistent, or* (ii) *consistent for a single value of k, in which case the α_i, α_{ij} are uniquely determined. In case* (i) *B_m is integral;*[3] *in case* (ii)

$$B_m = G_m - e \sum \frac{1}{P} \qquad (p^n \neq 2),^4$$

where G_m is integral, the summation is over all irreducible P of degree k, and

$$e = \frac{(-1)^{nk+k(\alpha_1+\cdots+s\alpha_s)}}{\prod_{ij} (\alpha_{ij}!)}.$$

The purpose of the present note is (1) to replace the conditions (1.1) by the following simpler conditions:

$$(1.2)\qquad \begin{aligned} m &= \sum_h \beta_h p^h & & (0 \leqq \beta_h < p), \\ nk(p-1) &= \sum_h \beta_h, & & p^{nk} - 1 \mid m; \end{aligned}$$

and (2) to simplify the original proof—more precisely the proof of Theorem 8 in that paper.

2. We first show that systems (1.1) and (1.2) are equivalent. If we assume that (1.1) is satisfied, it follows almost immediately that (1.2) holds. For put

$$(2.1)\qquad \beta_{nki+j} = \alpha_{ij} \qquad (0 \leqq j < nk);$$

then

$$\sum_h \beta_h p^h = \sum_{ij} \alpha_{ij} p^{nki+j} = \sum_i p^{nki} \sum_j \alpha_{ij} p^j = \sum_i p^{nki} \alpha_i = m;$$

$$\sum_h \beta_h = \sum_j \sum_i \alpha_{ij} = \sum_j (p-1) = nk(p-1);$$

$$m = \sum_i \alpha_i p^{nki} \equiv \sum_i \alpha_i \equiv 0 \pmod{p^{nk} - 1},$$

so that the several parts of (1.2) are verified.

[2] This Journal, vol. 3(1937), pp. 503–517.
[3] That is, a polynomial.
[4] For $p^n = 2$, see p. 517 of second reference.

Conversely, assume that (1.2) is satisfied. We may put

$$m = \sum_{i=0}^{s-1} \alpha_i p^{nki}, \qquad \alpha_i = \sum_{j=0}^{nk-1} \alpha_{ij} p^j, \qquad 0 \leqq \alpha_{ij} \leqq p - 1; \tag{2.2}$$

clearly (2.1) holds. From $p^{nk} - 1 \mid m$ and the first of (2.2) it follows that

$$p^{nk} - 1 \mid \sum \alpha_i . \tag{2.3}$$

Next write

$$\sum_i \alpha_i = \sum_{ij} \alpha_{ij} p^j = \sum_j \epsilon_j p^j, \tag{2.4}$$

where for brevity

$$\epsilon_j = \sum_i \alpha_{ij}.$$

Again by (2.1)

$$\sum_h \beta_h = \sum_{ij} \beta_{nki+j} = \sum_{ij} \alpha_{ij} = \sum_j \epsilon_j ,$$

and therefore, by one of the conditions in (1.2),

$$\sum_{j=0}^{nk-1} \epsilon_j = nk(p - 1). \tag{2.5}$$

From (2.3) and (2.4) follows

$$\sum_{j=0}^{nk-1} \epsilon_j p^j = \mu_0 (p^{nk} - 1), \tag{2.6}$$

where μ_0 is an integer > 0. Multiplying both members of (2.6) by p and eliminating p^{nk} on the left, we get

$$\epsilon_0 p + \cdots + \epsilon_{nk-2} p^{nk-1} + \epsilon_{nk-1} = \mu_1 (p^{nk} - 1),$$

where μ_1 is an integer > 0. Continuing in this way we get the following system of equations:

$$\begin{aligned} \epsilon_0 p^j + \cdots + \epsilon_{nk-j-1} p^{nk-1} + \epsilon_{nk-j} + \cdots + \epsilon_{nk-1} p^{j-1} &= \mu_j (p^{nk} - 1) \\ &(0 \leqq j \leqq nk - 1), \end{aligned} \tag{2.7}$$

where μ_j is an integer > 0. We shall show that $\mu_j = 1$. For adding corresponding members of (2.7) we get

$$\sum \epsilon_j \cdot \sum p^j = \sum \mu_j \cdot (p^{nk} - 1),$$

so that

$$\sum \epsilon_j = (p - 1) \sum \mu_j .$$

Comparison with (2.5) leads to

$$\sum_{j=0}^{nk-1} \mu_j = nk;$$

but since each $\mu_j \geqq 1$, it follows immediately that

$$\mu_j = 1 \qquad (0 \leqq j \leqq nk - 1). \tag{2.8}$$

Now using the j-th and $(j + 1)$-th equation in (2.7) we get

$$\epsilon_{nk-j-1} = p\mu_j - \mu_{j+1} = p - 1 \qquad (j < nk - 1),$$

by (2.8); finally $\epsilon_0 = p - 1$ in view of (2.5). Thus we have proved

$$\epsilon_j = \sum_i \alpha_{ij} = p - 1 \qquad (0 \leqq j \leqq nk - 1),$$

and therefore the last equation in (1.1) is verified.

Note that for given m the first equation in (1.2) uniquely determines β_h, and it is easy to check whether the remaining equations in (1.2) are satisfied. Incidentally it is evident from the second equation that if a k exists it is necessarily unique.

3. We employ the same notions that were used in the former paper. In particular we shall call the series

$$H = \sum_{m=0}^{\infty} \frac{A_m}{g_m} t^m,$$

where the A_m are integral, an H-series. Thus the sum, difference and product of two H-series is again an H-series. Let H_1 denote an H-series without constant term; we have the result that H_1^λ/g_λ is an H-series.

By the congruence

$$\sum \frac{A_m}{g_m} t^m \equiv \sum \frac{A_m'}{g_m} t^m \pmod P$$

we shall understand the system of congruences

$$A_m \equiv A_m' \pmod P \qquad (m = 0, 1, \cdots).$$

Then by the result quoted we have

$$H_1^{p^{nk}} \equiv 0 \pmod P,$$

for P irreducible of degree k; more generally we have

$$H_1^\lambda \equiv 0 \pmod P \qquad (\lambda \geqq p^{nk}), \tag{3.1}$$

since in this case g_λ is a multiple of F_k.

We are now able to give a simplified proof of the following lemma (which may be considered the principal point in the proof of the main theorem).

LEMMA. *For P irreducible of degree k,*

$$\psi^{p^{nk}-1}(t) \equiv \left(\sum \frac{(-1)^{ki}}{F_{ki}}\, t^{p^{nki}}\right)^{p^{nk}-1} \pmod{P}. \tag{3.2}$$

Put

$$\psi = \psi(t) = \sum_{j=0}^{k-1} (-1)^j \psi_j, \tag{3.3}$$

where

$$\psi_j = \sum_{i=0}^{\infty} (-1)^{ki} \frac{t^{p^{n(ki+j)}}}{F_{ki+j}}.$$

Then we have

$$\begin{aligned} \psi_j^{p^n} &= \sum (-1)^{ki} \frac{t^{p^{n(ki+j+1)}}}{F_{ki+j}^{p^n}} \\ &= \sum (-1)^{ki} \frac{t^{p^{n(ki+j+1)}}}{F_{ki+j+1}} [ki + j + 1] \\ &\equiv [j+1] \sum (-1)^{ki} \frac{t^{p^{n(ki+j+1)}}}{F_{ki+j+1}} \pmod{P}, \end{aligned}$$

where as above $[k] = x^{p^{nk}} - x$. Thus we see that

$$\psi_j^{p^n} \equiv [j+1]\psi_{j+1} \pmod{P} \tag{3.4}$$

for $j < k - 1$, while for $j = k - 1$ we have

$$\psi_{k-1}^{p^n} \equiv 0 \pmod{P}.$$

Now from (3.4) follows

$$\psi_0^{p^n} \equiv F_1\psi_1,$$

$$\psi_0^{p^{2n}} \equiv [2]\psi_1^{p^n} \equiv F_2\psi_2,$$

and generally

$$\psi_0^{p^{nj}} \equiv F_j\psi_j \pmod{P} \tag{3.5}$$

for $j < k$; for $j = k$ the congruence becomes

$$\psi_0^{p^{nk}} \equiv 0 \pmod{P}.$$

Since $F_j \not\equiv 0 \pmod{P}$ for $j < k$, we get from (3.3) and (3.5)

$$\psi \equiv \sum_{j=0}^{k-1} \frac{(-1)^j}{F_j}\, \psi_0^{p^{nj}} \pmod{P}. \tag{3.6}$$

Turning now to (3.2), put

$$\psi^{p^{nk}-1} = (\psi\psi^{p^n} \cdots \psi^{p^{n(k-1)}})^{p^n-1}$$

and substitute from (3.6). Then we get

$$
\begin{aligned}
\psi^{p^{nk}-1} &\equiv \left(\psi_0 - \frac{\psi_0^{p^n}}{F_1} + \cdots\right)^{p^n-1} (\psi_0^{p^n} - \cdots)^{p^n-1} \cdots \\
&\equiv \psi_0^{p^{nk}-1} + R,
\end{aligned} \tag{3.7}
$$

where R stands for the sum of the other terms obtained by expanding the product in (3.7). Clearly each term in R involves ψ_0^λ, where $\lambda \geqq p^{nk}$, and therefore by (3.1) each such term $\equiv 0 \pmod P$. We have therefore

$$\psi^{p^{nk}-1} \equiv \psi_0^{p^{nk}-1} \pmod P,$$

so that we have proved the lemma.

4. As an instance of the use of (1.2) we may easily show that *for given $k > 0$ there exist infinitely many m satisfying* (1.2) *and therefore also* (1.1). Indeed it suffices to put

$$\beta_h = p - 1 \qquad (0 \leqq h \leqq nk - 1),$$

$$m = \sum_{h=0}^{nk-1} \beta_h p^{h\lambda} = (p-1)\frac{p^{nk\lambda} - 1}{p^\lambda - 1}. \tag{4.1}$$

If now we take

$$\lambda \equiv 1 \pmod{nk}, \tag{4.2}$$

but otherwise arbitrary, then

$$
\begin{aligned}
m = (p-1)\sum_{h=0}^{nk-1} p^{h\lambda} &\equiv (p-1)\sum_{h=0}^{nk-1} p^h \\
&\equiv p^{nk} - 1 \equiv 0 \pmod{p^{nk} - 1};
\end{aligned}
$$

also

$$\sum_{h=0}^{nk-1} \beta_h = (p-1)\sum_{h=0}^{nk-1} 1 = nk(p-1),$$

so that (1.2) is verified.

From the above we have the following corollary to the main theorem:

For fixed $k > 0$ there are infinitely many B_m with fractional part

$$\sum_{\deg P = k} \frac{1}{P}.$$

In general it suffices to take m as given by (4.1) and (4.2). However, for $p^n = k = 2$, we take

$$m = 2^{\lambda+1} + 2, \qquad \lambda \equiv 1 \pmod 2.$$

DUKE UNIVERSITY.

ARC-SPACES

By J. W. T. Youngs

Introduction

In the study of the structure of a continuous curve, G. T. Whyburn introduced and widely applied the notion of a cyclic element.[1] One of the interesting results of the theory is that, if each cyclic element of a continuous curve is considered as a point of a hyperspace (that is, we may say, if the continuous curve is reduced modulo cyclic elements), then this hyperspace displays a dendritic structure. This phrase is not usually intended to imply that the hyperspace is a space in any well-defined sense. In fact, it serves only to draw attention to the similarity between cyclic chains and arcs, and above all to the property that the cyclic chain between any two points is unique. As a matter of fact, the hyperspace usually bears no resemblance to a Peano space. In this the situation is quite different from that encountered when we have an upper semi-continuous partition of a Peano space; for, if we reduce the Peano space modulo the elements of the partition, the resulting collection can be topologized so as to be a Peano space again.

An example will serve to illustrate the point. Consider two tangent circles in the xy-plane. Here we have exactly three cyclic elements (one degenerate and two true) so that the hyperspace consists of three points. What is the meaning of the terms space and arc when applied to a situation of this sort?

One sense in which we may speak of the collection of cyclic elements as a space, and more particularly a dendrite, follows as a consequence of some general results obtained by R. L. Moore.[2] More recently, some unpublished results of A. D. Wallace might be applied to this situation.

It may be said that our main purpose is to consider the question outlined above. In doing so a study will be made of spaces (known as arc-spaces) satisfying a certain set of axioms in terms of the notions of point and arc. We shall attempt to parallel the essential features of the work of Whyburn on Peano spaces so as not to have to go into detail where the proofs are of a standard character.

Received July 21, 1939; in revised form, July 27, 1940. This paper was begun while the author was in residence at Charlottesville. He wishes to take this opportunity to thank Professor G. T. Whyburn for offering him the facilities of the Department of Mathematics at the University of Virginia.

[1] For a bibliography of the literature up to the year 1930, see C. Kuratowski and G. T. Whyburn, *Sur les éléments cycliques et leurs applications*, Fund. Math., vol. 16(1930), pp. 330–331; and also G. T. Whyburn, *On the structure of continua*, Bull. Amer. Math. Soc., vol. 42(1936), pp. 71–73.

[2] R. L. Moore, *Fundamental Theorems Concerning Point Sets*, Rice Institute Pamphlets, vol. 23, no. 1, 1936; especially p. 74.

An arc-space will be found to display a definite structure in terms of its cyclic elements, and when these are considered as points of a hyperspace, the hyperspace can be "topologized" so as to be an arc-space. In fact, the hyperspace will be found to be a dendrite in the sense that there is one and only one arc joining any two points of the hyperspace. The hyperspace of the hyperspace will be a dendrite and a repetition of the process yields no essentially different space. We shall proceed to develop these remarks by considering first the well-known concept of ordered set.

I. Preliminaries

1.1. Suppose that α denotes a set of distinct elements $x, y, z, \cdots$. The set α is said to be *ordered* if there is a binary relation R defined for *distinct* elements of the set α such that: (1) If x and y are distinct elements of α, then either $x\ R\ y$ or $y\ R\ x$. (2) If $x\ R\ y$ and $y\ R\ z$, then $x\ R\ z$. It follows that $x\ R\ y$ excludes $y\ R\ x$; for, otherwise $x\ R\ x$ by (2) and R is defined only for distinct elements.

1.2. If $x\ R\ y$, then x is said to *precede* y and y is said to *follow* x. An element of α which is not preceded by any is called the *first* element of α; one which is not followed by any is called the *last*. Any ordered set in which we are interested will have both a first and a last element. If a is the first and b the last element of an ordered set α, then by $\alpha(a, b)$ will be understood the system consisting of the set α together with the ordering relation R.

1.3. It is convenient to replace the notation $x\ R\ y$ by the equivalent form $x < y\ [\alpha(a, b)]$, or $y > x\ [\alpha(a, b)]$. One reason for doing this is the following. Another binary relation R^{-1} may be defined in α as follows: $x\ R^{-1}\ y$ means $y\ R\ x$. This is called the *reverse order* and now b is the first while a is the last element of the set. This newly ordered set is denoted by $\alpha(b, a)$ and the notation $x\ R^{-1}\ y$ is replaced by $x < y\ [\alpha(b, a)]$, or $y > x\ [\alpha(b, a)]$.

1.4. If it is necessary to consider the set of elements in $\alpha(a, b)$ without any reference to their order, then the comma in the notation is omitted and the symbol $\alpha(ab)$ is used. Thus $\alpha(ab) = \alpha(ba)$ because they consist of the same elements. As a rule this care to distinguish between the ordered set $\alpha(a, b)$ and the set $\alpha(ab)$ will not be necessary.

1.5. Two ordered sets $\alpha_1(a_1, b_1)$ and $\alpha_2(a_2, b_2)$ are said to be the *same* (notation: $\alpha_1(a_1, b_1) = \alpha_2(a_2, b_2)$) if and only if (1) $\alpha_1(a_1b_1) = \alpha_2(a_2b_2)$, (2) $x < y$ $[\alpha_1(a_1, b_1)]$ implies $x < y\ [\alpha_2(a_2, b_2)]$.

1.6. If $\alpha(a, b)$ is an ordered set, and $x < y\ [\alpha(a, b)]$, then by the *segment* $\alpha(x, y)$ of the ordered set $\alpha(a, b)$ we shall mean the ordered set whose elements consist of all z's such that $x \leqq z \leqq y\ [\alpha(a, b)]$, and whose order is defined by the convention that $u < v\ [\alpha(x, y)]$ if and only if $u < v\ [\alpha(a, b)]$.[3]

[3] By $x \leqq y\ [\alpha(a, b)]$ we mean that either $x = y$ or $x < y\ [\alpha(a, b)]$.

1.7. If $\alpha_1(a, x)$ and $\alpha_2(x, b)$ are ordered sets having only x in common, then their *sum* $[\alpha_1(a, x) + \alpha_2(x, b)]$ is understood to be the ordered set $\alpha(a, b)$ whose elements are $[\alpha_1(ax) + \alpha_2(xb)]$ and whose order is defined as follows: $u < v$ $[\alpha(a, b)]$ if and only if $u \neq v$ and in addition one of these three conditions is fulfilled: (1) $u < v$ $[\alpha_1(a, x)]$, (2) $u < v$ $[\alpha_2(x, b)]$, (3) $u \in \alpha_1(ax)$, $v \in \alpha_2(xb)$.

1.8. If E is any subset of the set $\alpha(ab)$, we say E has an *order relative to* $\alpha(a, b)$. In fact, if $x, y \in E$, then $x\ R\ y$ if and only if $x < y$ $[\alpha(a, b)]$. We can speak of a first and last element of E along $\alpha(a, b)$ if such elements exist. The ordered set $\alpha(a, b)$ is said to be *well ordered* if each of its non-null subsets has a first element. Our principal interest will be a property which may be called a modification of this. Suppose that we are given a class K of ordered sets whose elements are chosen from a set 1. The class K is said to be *proper* if for every $\alpha_1(a, b)$, $\alpha_2(c, b) \in K$, the set of points common to both has a first element along $\alpha_1(a, b)$.

1.9. Mention should be made of what is known as a *cut* of $\alpha(a, b)$. A cut $[L, U]$ of $\alpha(a, b)$ is a classification of all the elements of $\alpha(a, b)$ into two non-vacuous sets L and U such that if $l \in L$ and $u \in U$, then $l < u$ $[\alpha(a, b)]$. A cut is said to be *localized* unless there is no last element in L and no first element in U.

II. Arc and arc-space

2.1. Let the symbol 1 stand for a set of elements (called points) which will be denoted by the small Latin letters $a, b, c, \cdots$. Suppose that there is given a class K of ordered subsets $\alpha(a, b)$ composed of two or more distinct elements of 1. Suppose further that each $\alpha(a, b)$ contains a first and last element which we shall denote generically by a and b respectively.

2.2. It is agreed that any ordered set $\alpha(a, b) \in K$ will be called an *arc from a to b*, while the class K will be called the class of arcs in 1. The term arc will be applied to an ordered set of elements in 1 only if it is in K.

2.3. If $x < y$ $[\alpha(a, b)]$, $\alpha_1(x, y) \in K$, $\alpha_1(xy) \subset \alpha(ab)$, then the arc $\alpha_1(x, y)$ is called a *subarc of* $\alpha(a, b)$ *from* x *to* y. The points a and b of the set $\alpha(ab)$ are known as *end points* of the arc $\alpha(a, b)$. The notation $\alpha(ab)^0$ is used to mean the set $[\alpha(ab) - (a + b)]$ which is called the *interior of the arc* $\alpha(a, b)$. If $\alpha(ab)^0 = 0$, then a is said to be *adjacent* to b and the relationship is denoted by a/b.[4]

2.4. If the set of elements of 1 together with the class K satisfy the following axioms, then the system consisting of points and arcs is known as an *arc-space*.

A1. *Symmetry of arcs.* If $\alpha(a, b) \in K$, then $\alpha(b, a) \in K$.

A2. *Existence and uniqueness of subarcs.* If $x < y$ $[\alpha(a, b)]$, then $\alpha(x, y) \in K$ and is the only subarc of $\alpha(a, b)$ from x to y. (See §§1.6 and 2.3.)[5]

[4] The concept is similar to the primitive concept of contiguity used by R. L. Moore, loc. cit., p. 1.

[5] By $\alpha(x, y)$ we understand a subarc of $\alpha(a, b)$. If we wish to deal with another arc from x to y, the notation $\alpha_1(x, y)$, $\alpha'(x, y)$, etc. will be used.

A3. *Completeness of space.* Every cut is localized. (See §1.9.)

A4. *Finite additivity of arcs.* If $\alpha_1(a, x)$ and $\alpha_2(x, b)$ are arcs and have only x in common, then their sum is an arc. (See §1.7.)

A5. *The set K is proper.* (See §1.8.)

A6. *Connectivity of space.* Any two points of 1 are end points of some arc.

III. Some consequences of the axioms

The axioms are clearly abstracted with the properties of ordinary arcs in mind. Some of the more immediate consequences are presented below.

3.1. THEOREM. *No proper subset of an arc can constitute an arc with the same end points.*

The proof is made by using A2.

3.2. THEOREM. *If two arcs $\alpha_1(a, b)$ and $\alpha_2(x, y)$ have the property that $\alpha_1(ab) = \alpha_2(xy)$ but $\alpha_1(a, b) \neq \alpha_2(x, y)$, then $\alpha_2(x, y) = \alpha_1(b, a)$.*

Proof. First suppose that $b \leqq x < y \leqq a$ $[\alpha_1(b, a)]$. By A2, $\alpha_1(x, y)$ is a subarc of $\alpha_1(b, a)$ from x to y. But $\alpha_1(xy) \subset \alpha_1(ba) = \alpha_2(xy)$, hence by §3.1, $\alpha_1(x, y) = \alpha_2(x, y)$. It follows that $\alpha_1(b, a) = \alpha_2(x, y)$. If $a \leqq x < y \leqq b$ $[\alpha(a, b)]$, a similar argument shows that $\alpha_2(x, y) = \alpha_1(a, b)$, and this is ruled out by hypothesis.

3.3. If there are two arcs $\alpha_1(a, x)$ and $\alpha_2(x, b)$ such that they have only x in common, then their sum is an arc $\alpha(a, b)$ by A4. Now $x \in \alpha(ab)$ and so there are unique subarcs $\alpha(a, x)$ and $\alpha(x, b)$. Using A2, we easily see that $\alpha_1(a, x) = \alpha(a, x)$ and $\alpha_2(x, b) = \alpha(x, b)$.

3.4. By definition, a/b if and only if there is an arc $\alpha(a, b)$ such that $\alpha(ab)^0 = 0$. A priori, there is no reason to suppose that there should not be another arc, say $\alpha_1(a, b)$ such that $\alpha_1(ab)^0 \neq 0$. Because of A2 we are able to prove the often applied

THEOREM. *If a/b, then there is no arc $\alpha(a, b)$ such that $\alpha(ab)^0 \neq 0$.*

Proof. If there were such an arc $\alpha(a, b)$, then $a + b \subset \alpha(ab)$. But because a/b, a and b in this order constitute an arc from a to b, and this is contradictory to §3.1.

COROLLARY. *If a/b and b/c, then a/c is false.*

Proof. If a/c, then a and c in this order form an arc from a to c. But by A4 there is an arc $\alpha(a, c)$ consisting of a, b and c in this order.

3.5. A point x is said to be *adjacent to y along $\alpha(a, b)$* if $x < y$ $[\alpha(a, b)]$ and x/y. The reader will readily be able to construct examples in which a point may be adjacent to any number of points. One such is the following. Let the points of the space consist of the points on the circumference of the unit circle and the center o. Let a and b represent points on the circumference. Ordered sets of the following type make up the class K. (1) a, o, b. (2) o, a. (3) a, o.

It is easily verified that this is an arc-space with every point on the circumference adjacent to the center.[6]

In particular, however, we have the following easily proved

THEOREM. *If x/y along $\alpha(a, b)$, then x is not adjacent to any other point along $\alpha(a, b)$.*

3.6. In the study of Peano spaces an important property of arcs is that if an arc from a to b and an arc from d to c have a point in common, then unless a is the same point as c, there is an arc from a to c. A similar situation exists in arc-spaces.

THEOREM. *Given $\alpha_1(a, x)$, $\alpha_2(x, b)$ and $a \neq b$, there is an arc $\alpha(a, b)$ such that $\alpha(ab) \subset [\alpha_1(ax) + \alpha_2(xb)]$.*

Proof. Using A5, suppose that z is the first point of $\alpha_1(a, x)$ which belongs to $\alpha_2(x, b)$. If $a \neq z \neq b$, then $\alpha_2(z, b)$ has no point other than z on $\alpha_1(a, z)$. By A4, $\alpha(a, b) = [\alpha_1(a, z) + \alpha_2(z, b)]$ is an arc. If $z = a$ or b, the theorem is obvious.

COROLLARY. *If an arc $\alpha(a, b)$ guaranteed by the hypothesis of the above theorem is such that $\alpha(ab)^0 = 0$, then a/b along $\alpha_1(a, x)$ or b/a along $\alpha_2(b, x)$, not both.*

Proof. Suppose that a/b is false along $\alpha_1(a, x)$ and b/a is false along $\alpha_2(b, x)$. Then using the method of the theorem we get an arc $\alpha'(a, b)$ which has at least one interior point. This is impossible by §3.4. On the other hand, if a/b along $\alpha_1(a, x)$ and b/a along $\alpha_2(b, x)$, consider $\alpha_1(b, x)$ and $\alpha_2(a, x)$. It is true that $a \epsilon' \alpha_1(bx)$ and $b \epsilon' \alpha_2(ax)$; hence the method of the theorem gives an arc $\alpha'(a, b)$ which must have at least one interior point. This is again impossible.

IV. Examples of arc-spaces

One example of an arc-space has already been given. Perhaps the most natural example is that of a Peano space with the ordinary interpretation of arc. In this section attention will be fixed on the independence of the axioms and a few more examples of arc-spaces.

4.1. We shall exhibit a space satisfying all the axioms except A5. The reader will have no difficulty in constructing similar examples to show independence on the part of the other axioms.

In the xy-plane consider the set of points $(0, -1)$, $(0, 1)$, $(x, 0)$ for $0 < x \leqq 1$. This set of points will constitute the points of the space. We shall completely describe the class K if we describe all the arcs between any two points as end points. If a and b are on the x-axis, then $\alpha(a, b)$ is unique and is the segment of the x-axis from a to b with the ordering taken from the axis itself. If $a = (0, -1)$, $b = (0, 1)$, then $\alpha(a, b) = a, b$. To conform with A1, $\alpha(b, a) = b, a$. In case $a = (0, -1)$, and b is a point $(x_1, 0)$ on the x-axis, $\alpha(a, b)$ consists of the

[6] This also gives an example of different arcs which have the same interior.

set $a + \underset{x}{E}[0 < x \leqq x_1]$, where a is the first point and the ordering of the other points is taken from the x-axis in increasing order of magnitude. $\alpha(b, a)$ consists of the reverse order. If $a = (0, 1)$ and b is on the x-axis, a similar definition of $\alpha(a, b)$ is adopted. But now the arc from $(0, 1)$ to $(1, 0)$ has no first point in common with the arc from $(0, -1)$ to $(1, 0)$, though the other axioms are satisfied.

4.2. The example of this section will later serve as an explanation for the unusual definition of a true cyclic element.

The points of the space are those in the xy-plane which satisfy the inequality $1 \leqq x^2 + y^2 \leqq 4$. If a and b are on the same ray, then the segment of this ray joining them is the unique arc $\alpha(a, b)$. If a and b are both on the circle $x^2 + y^2 = 1$, then the arc of this circle joining a to b in the counterclockwise direction is an arc $\alpha(a, b)$. That joining a to b in the clockwise direction is another. Suppose a is on the inner circle but b is not. If x is the point where the ray containing b cuts the inner circle, then we have previously defined a unique $\alpha(x, b)$ and two possible arcs $\alpha_1(a, x)$, $\alpha_2(a, x)$. The ordered set $[\alpha_i(a, x) + \alpha(x, b)]$ is an arc $\alpha_i(a, b)$ for $i = 1$ or 2. (See §1.7.) The reverses of these ordered sets constitute arcs from b to a. If neither a nor b is on the inner circle, or on the same ray, then we define the arcs $\alpha_i(a, b)$ for $i = 1, 2$ similarly using three steps and always employing segments of rays to get to the inner circle. It is not difficult to see that this is an arc-space.

4.3. For our final example the points of the space are the points (x, y) of the open square $0 < x < 1$, $0 < y < 1$ and the origin. If a and b are in the open square, any ordinary arc in the open square is an arc $\alpha(a, b)$. If a is the origin and b is any point in the open square, an arc $\alpha(a, b)$ consists of an arc in the ordinary sense, which, however, must have an initial segment from a lying on the line $y = x$. The reverse of such an arc is an arc from b to a. It is again not difficult to see that this is an arc-space.

V. Connected sets

5.1. A set E is said to be *connected* if $a, b \epsilon E$ and $a \neq b$ imply that there exists an arc $\alpha(a, b) \subset E$. In particular, if E consists of a single point, then E is connected.

Using the theorem of §3.6, we easily prove the following

THEOREM. *If Γ is a collection of connected sets and there is a connected set E such that every set $G \epsilon \Gamma$ has a point in common with E, then $E + \sum G$ is connected, where the summation is taken over all the sets G in Γ.*

5.2. If $a \epsilon E$, the *component* S_a of E containing a is the set consisting of a and all points x such that there exists an arc $\alpha(a, x) \subset E$. It is a simple matter to prove the following

THEOREM. *If $S_a \cdot S_b \neq 0$, then $S_a = S_b$.*

THEOREM. *S_a is the maximal connected subset of E containing a.*

5.3. A point x is said to *cut the space* (set E) *between a and b* if $x \neq a, b$ and the components S_a and S_b of $1 - x$ $(E - x)$ are distinct. A point x is said to cut the space (set E), or is said to be a cut point of the space (set E), if it cuts the space (set E) between two points.

THEOREM. *If x cuts the space between a and b and $\alpha(a, b)$ is any arc from a to b, then $x \in \alpha(ab)$.*

Proof. If the theorem is false, then there is an arc $\alpha(a, b)$ such that $x \in' \alpha(ab)$; i.e., $\alpha(ab) \subset 1 - x$, and so $S_a = S_b$ by §§5.1 and 5.2. But then x does not cut the space between a and b.

5.4. THEOREM. *If every arc $\alpha(a, b)$ between a and b is such that it passes through a fixed point $x \neq a, b$, then x cuts the space between a and b.*

The proof is immediate.

5.5. THEOREM. *If we have z_1/z_2 along $\alpha(a, b)$* (see §3.5), *then for any $\alpha'(a, b)$ it is true that z_1/z_2 along $\alpha'(a, b)$. In particular, if $z_i \neq a, b$, then z_i cuts the space between a and b for $i = 1, 2$.*

It is easy to see that a contradiction of the theorem leads to a contradiction of §3.4.

5.6. THEOREM. *If $y \in \alpha(ab)^0$ does not cut the space between a and b but there is at least one point which follows y on $\alpha(a, b)$ and does cut the space between a and b, then there is a first such point.*

Proof. Let $x \in L$ if $x \leqq y$ $[\alpha(a, b)]$, or if, for $x > y$ $[\alpha(a, b)]$, $\alpha(y, x)$ contains no points which cut the space between a and b. Let $x \in U$ otherwise. By A3 we get a point z such that $w > z$ $[\alpha(a, b)]$ implies that $\alpha(z, w)$ contains a point cutting the space between a and b, while $y < w < z$ $[\alpha(a, b)]$ implies that $\alpha(y, w)$ contains no such point. Consider any $\alpha_1(a, b)$. If we use A5 on $\alpha(z, b)$ and $\alpha_1(a, b)$, the above property of z in conjunction with §5.3, and §5.5 if necessary, it follows that $z \in \alpha_1(ab)$. Therefore, by §5.4, z cuts the space between a and b. Hence $y < z$ $[\alpha(a, b)]$ and z is clearly the first cut point following y.

5.7. DEFINITION. a is said to be *conjugate* to b (notation: $a \sim b$) if for any $x \neq a, b$ the components S_a and S_b of $1 - x$ are identical.[7]

REMARK. If a/b, then $a \sim b$.

5.8. A point x is said to be an *end point of the space* if it is an end point of every arc which contains it.

5.9. THEOREM. *If a point is conjugate to no point and is not a cut point, then it is an end point.*

Proof. Let y be the point. If the theorem is false, then there exists an arc $\alpha(a, b)$ which contains the point y and $y \neq a, b$. As y is not a cut point, $y \sim b$

[7] See Kuratowski and Whyburn, loc. cit., p. 306.

or it is conjugate to the first cut point between y and b. (See §5.6.) This is contrary to hypothesis.

We cannot assert that an end point is not conjugate to any point as it might be adjacent to some point.

5.10. The relationship of conjugacy is obviously reflexive and symmetric, though not transitive. However, we do have the following

THEOREM. *If $a \sim x \sim b$ and y cuts the space between a and b, then $x = y$.*

Proof. If $y \neq x$ consider $1 - y$. Now $a \sim x$, hence $S_a = S_x$. But $x \sim b$ and so $S_x = S_b$. Therefore $S_a = S_b$, and this is contrary to hypothesis.

COROLLARY.[8] *If $a \sim x \sim b$ and $a \sim y \sim b$, $x \neq y$, then $a \sim b$.*

VI. Cyclic elements

If we examine the example of §4.2, we see that every point, with the exception of those on the outer circle, is a cut point. If we define a true cyclic element as the totality of points conjugate to a non-cut point which is not an end point, then in this case we find there are no true cyclic elements.[9] It seems desirable, however, for the inner circle of our space to be considered as a true cyclic element. The definition adopted brings this about and is at the same time equivalent to the above cited definition in a Peano space.

6.1. A *true cyclic element* is the totality of points conjugate to each of two distinct points which are conjugate to each other. That is, if $a \neq b$, and $a \sim b$, then they generate a true cyclic element $M(a, b)$ defined by the rule that $x \in M(a, b)$ if and only if $a \sim x \sim b$.[10]

A *degenerate cyclic element* is a cut point or an end point.

6.2. THEOREM. *If $x, y \in M(a, b)$, and $x \neq y$, then $x \sim y$ and $M(a, b) = M(x, y)$.*

Proof. $x \sim a \sim y$ and $x \sim b \sim y$, therefore $x \sim y$ by the corollary to §5.10. Hence the notation $M(x, y)$ is justified. If $z \in M(a, b)$, then $z \sim a \sim x$, $z \sim b \sim x$, and so $z \sim x$. Similarly, $z \sim y$. Therefore $z \in M(x, y)$. Hence $M(a, b) \subset M(x, y)$. Similarly, $M(x, y) \subset M(a, b)$.

We shall conclude by stating a sequence of theorems which follow in as simple a fashion as the above.

6.3. THEOREM. *If $M(a, b) \cdot M(c, d) \supset x + y$ $(x \neq y)$, then $M(a, b) = M(c, d)$.*

6.4. THEOREM. *If $M(a, b) \cdot M(c, d) = x$, then x cuts the space.*

6.5. THEOREM. *The cyclic elements cover the space.*

[8] A treatment of cyclic elements using this weak transitivity was given by T. Radó at the April meeting of the American Mathematical Society in Chicago, 1939.

[9] See Kuratowski and Whyburn, loc. cit., p. 306.

[10] In a Peano space this definition of a cyclic element is equivalent to that of Kuratowski and Whyburn, loc. cit., p. 306.

VII. Sets of type $\mathcal{H}$

7.1. A set H is of type $\mathcal{H}$ if it fulfills the following condition:

(ϕ): *If a and b are two distinct points in E, then every arc $\alpha(a, b)$ is contained in E.*[11]

The empty set fulfills condition (ϕ) vacuously, as does any set consisting of but a single point.

7.2. THEOREM. *If Z is a connected set and H is of type $\mathcal{H}$, then $Z \cdot H$ is connected.*

7.3. THEOREM. *If x cuts H, a set of type $\mathcal{H}$, between a and b, then x cuts the space between a and b.*

7.4. It will be noticed that we deviate somewhat from the usual procedure in the study of cyclic elements in a Peano space by considering sets of type $\mathcal{H}$ instead of sets of type $\mathcal{A}$. This difference is unavoidable as we have no concept of closure in an arc-space. The frontier point of a component of the complement of a set of type $\mathcal{A}$ played an important rôle in the cyclic element theory. To get to the analogous situation we proceed in the following way.

If $a \epsilon H$, $b \epsilon S_b$ (where, unless otherwise stated, H is of type $\mathcal{H}$ and S_b is the component of $1 - H$ containing b) and $\alpha(a, b)$ is any arc from a to b, then a point z is said to have property (f) on $\alpha(a, b)$ if $z \epsilon \alpha(ab)$ and $x > z\ [\alpha(a, b)]$ implies $x \epsilon S_b$, while $x < z\ [\alpha(a, b)]$ implies $x \epsilon H$.

THEOREM. *If $a \epsilon H$ and $b \epsilon S_b$, then on each $\alpha(a, b)$ there is a point z having property (f) on $\alpha(a, b)$.*

Proof. Let $x \epsilon L$ if $x = a$, or $\alpha(ax) \subset H$. Otherwise $x \epsilon U$. By A3 there is a z such that if $x < z\ [\alpha(a, b)]$, then $\alpha(ax) \subset H$; and if $x > z\ [\alpha(a, b)]$, then $\alpha(ax) \not\subset H$. In the first case $x \epsilon H$. In the second, suppose that $x \epsilon' S_b$. In the event $x \epsilon H$ property (ϕ) is contradicted as $\alpha(ax) \not\subset H$. Therefore $x \epsilon' H$ and so $x \epsilon S_x$. But $\alpha(x, b)$ joins x of S_x to b of S_b in $1 - H$. This means $S_x = S_b$. Therefore $x \epsilon S_b$ and the theorem is proved.

COROLLARY. *The point z defined above must belong to H or to S_b.*

7.5. THEOREM. *If $a \epsilon H$, $b \epsilon S_b$ and on a certain $\alpha(a, b)$ there are two points $z_1 < z_2\ [\alpha(a, b)]$ having property (f), then z_1/z_2 and $z_1 \epsilon H$, $z_2 \epsilon S_b$.*

Proof. $\alpha(z_1, z_2) = z_1, z_2$; for if there were a point $w \epsilon \alpha(z_1z_2)^0$, then $w < z_2\ [\alpha(a, b)]$ and so $w \epsilon H$, but $w > z_1\ [\alpha(a, b)]$, hence $w \epsilon S_b$. Since $z_1 < z_2\ [\alpha(a, b)]$, $z_1 \epsilon H$ and $z_2 \epsilon S_b$.

COROLLARY. *There cannot be three points with property (f) on $\alpha(a, b)$.*

7.6. THEOREM. *If $z_i \epsilon \alpha_1(ab)$, $z_i \epsilon \alpha_2(ab)$, and z_i has property (f) on $\alpha_i(a, b)$ for $i = 1, 2$, then z_i has property (f) on $\alpha_{3-i}(a, b)$ for $i = 1, 2$.*

Proof. Suppose $a \neq z_i \neq b$ for $i = 1, 2$, and $z_1 < z_2\ [\alpha_1(a, b)]$. Then $z_1 < z_2\ [\alpha_2(a, b)]$; for if $z_2 < z_1\ [\alpha_2(a, b)]$, then $z_2 \epsilon H$, but $z_1 < z_2\ [\alpha_1(a, b)]$ implies $z_2 \epsilon S_b$.

[11] See Kuratowski and Whyburn, loc. cit., p. 317 and p. 321.

Now $\alpha_1(z_1z_2)^0 \subset S_b$, and $\alpha_2(z_1z_2)^0 \subset H$. It is a simple matter to show that $\alpha_2(z_1z_2)^0 = 0$ and so both are empty, for otherwise property (ϕ) is contradicted. The theorem now follows. The other cases may be handled similarly.

7.7. THEOREM. *If z has property (f) on $\alpha(a, b)$ and $a_1 \in H$, $b_1 \in S_b$, then there exists an arc $\alpha'(a_1, b_1)$ such that z has property (f) on $\alpha'(a_1, b_1)$.*

Proof. Let the points be distinct. Take any $\alpha_1(a_1, a) \subset H$ and $\alpha_2(b_1, b) \subset S_b$. Now $\alpha_1(a_1, a)$ and $\alpha(a, z)$ contain an arc $\alpha_3(a_1, z) \subset H + z$, by §3.6, while $\alpha_2(b_1, b)$ and $\alpha(b, z)$ contain an arc $\alpha_4(b_1, z) \subset S + z$. But by A4, $[\alpha_3(a_1, z) + \alpha_4(z, b_1)]$ forms an arc $\alpha'(a_1, b_1)$ and z has property (f) on it. The proof is similar in the other cases.

7.8. THEOREM. *If z_1 has property (f) on $\alpha_1(a, b)$ and z_2 has property (f) on $\alpha_2(a, b)$, then $z_i \in \alpha_{3-i}(ab)$ and z_i has property (f) on $\alpha_{3-i}(a, b)$, for $i = 1, 2$.*

Proof. Suppose $a \neq z_i \neq b$ ($i = 1, 2$). Now if $z_2 \notin \alpha_1(ab)$, consider $\alpha_2(z_2, b)$ and $\alpha_1(z_1, b)$. We have an arc $\alpha(z_2, z_1) \subset \alpha_2(z_2, b) + \alpha_1(z_1, b)$ by §3.6. If $\alpha(z_2z_1)^0 = 0$, then $\alpha_1(a, b)$ contains z_2 and z_1, by §§3.7 and 5.5. Hence $0 \neq \alpha(z_2z_1)^0 \subset S_b$. Similarly, $\alpha_2(z_2, a)$ and $\alpha_1(z_1, a)$ contain an arc $\alpha'(z_2, z_1)$ such that $0 \neq \alpha'(z_2z_1)^0 \subset H$.

Now $\alpha(z_2, z_1)$ and $\alpha'(z_2, z_1)$ have only their end points in common, and it follows by the use of §3.4 that $\alpha'(z_2z_1)$ must contain at least two points. But this contradicts property (ϕ). Hence $z_2 \in \alpha_1(ab)$. Similarly, $z_1 \in \alpha_2(ab)$. By §7.6 it follows that z_i has property (f) on $\alpha_{3-i}(a, b)$ for $i = 1, 2$. The proof is similar in the other cases.

COROLLARY. *If z has property (f) on $\alpha(a, b)$, then z has property (f) on any $\alpha(a, b)$, and z cuts the space between a and b if $z \neq a, b$.*

COROLLARY. *If z has property (f) on $\alpha(a, b)$ and $a_1 \in H$, $b_1 \in S_b$, then z has property (f) on any $\alpha(a_1, b_1)$.*

7.9. The results of the above section show that if we are given a set H of type $\mathcal{H}$ and a component S of the complement of H, then a point z having property (f) along a certain arc $\alpha(a, b)$ with one end point in H and the other in S, will also have property (f) along any arc with one end point in H and the other in S. In other words, while the definition of a point z having property (f) along $\alpha(a, b)$ implied that the property was relative to the arc, we now see that it is really a property associated with the sets H and S. A point z of the above character will in the future be called a *frontier point between H and S*. In this new terminology we see that (1) there exists a frontier point between H and S, (2) it must belong to H or to S, (3) there can be at most two frontier points, (4) in case there are two, one belongs to H while the other belongs to S, and the two are adjacent.

The two theorems below follow immediately from the above discussion.

7.10. THEOREM. *If z is a frontier point between H and S, and $a \in H$, $b \in S$, $a, b \neq z$, then z cuts the space between a and b.*

7.11. Definition. If $z \epsilon S$ is the only frontier point between H and S, then z is known as a *stray frontier point* between H and S. By H^* we shall understand the set H together with all of its stray frontier points.

Theorem. *If z is a stray frontier point between H and S, then it is not adjacent to any point of H.*

7.12. Lemma. *If $x \epsilon H$ and S is a component of $H - x$, then $\alpha(a, x) \subset S + x$ for any arc $\alpha(a, x)$ if $a \epsilon S$.*

Proof. If S is degenerate, $S = a$ and therefore $\alpha(ax) = a + x \subset S + x$. If S is non-degenerate, suppose $y \epsilon \alpha(ax)^0$. If there is no such y, the theorem is obvious. Consider $\alpha(a, y)$. Since H is of type $\mathcal{H}$, $\alpha(ay) \subset H$, and as $x \epsilon' \alpha(ay)$, $\alpha(ay) \subset H - x$; i.e., $\alpha(ay) \subset S$. Therefore $\alpha(ax) \subset S + x$.

7.13. Theorem. *If $x \epsilon H$ and S is a component of $H - x$, then $S + x$ is of type H.*

Proof. If S is degenerate, the theorem is obvious. If S is non-degenerate, take two distinct points $a, b \epsilon (S + x)$, and consider any $\alpha(a, b)$. Now $\alpha(ab) \subset H$, and if $x \epsilon' \alpha(ab)$, then $\alpha(ab) \subset S$. If $x \epsilon \alpha(ab)$, then $\alpha(ax) \subset (S + x)$ and $\alpha(xb) \subset (S + x)$ by §7.12. Therefore, $\alpha(ab) \subset (S + x)$ and the theorem is proved.

7.14. For a set H of type $\mathcal{H}$ the operation of adding the stray frontier points is quite similar to the operation of taking the closure of a set in ordinary topological spaces. One difference is that whereas $\bar{E}$ has a meaning in a topological space for every set E, the set H^* is defined only if H is of type $\mathcal{H}$.

The reader will have no difficulty in showing that if H_1 and H_2 are of type $\mathcal{H}$, and $H_1 \cdot H_2 \neq 0$, then $H_1 + H_2$ is of type $\mathcal{H}$. From this it follows that $(H_1 + H_2)^* = H_1^* + H_2^*$. It is also true that $0^* = 0$ and $p^* = p$ for any $p \epsilon 1$. Finally, H^* is of type $\mathcal{H}$ and in fact $(H^*)^* = H^*$. This leads one to hope for an extension of the operation H^* to sets which are not necessarily of type $\mathcal{H}$, an extension which will make the operation formally identical to the operation of taking the closure of a set.

In this connection a result which will be applied later is that if $H_1 \supset H_2$, then $H_1^* \supset H_2^*$.

VIII. **Characteristic properties of cyclic elements**

By the *cyclic chain* from a to b ($a \neq b$) we mean the set $C(a, b) = \prod H$, where the product is taken over all sets H of type $\mathcal{H}$ containing a and b.[12]

It is evident that for any two points a and b there is one and only one cyclic chain $C(a, b)$, and it is a set of type $\mathcal{H}$. The meaning of $C^*(a, b)$ is clear from §7.11.

[12] See Kuratowski and Whyburn, loc. cit., p. 311. For Peano spaces the definitions are equivalent.

8.1. THEOREM. *If $a \sim b$ and $a \neq b$, then there are no cut points of $C(a, b)$.*

Proof. Suppose $x \in C(a, b)$, $x \neq a$, but possibly $x = b$. Let S_a be the component of $C(a, b) - x$ containing a. Now $(S_a + x)$ is a set of type $\mathcal{H}$ by §7.13, and as $a \sim b$ it follows by §7.3 that $b \in (S_a + x)$. Therefore, $(S_a + x) \supset C(a, b)$ by definition. But $(S_a + x) \subset C(a, b)$. Hence $S_a = C(a, b) - x$ and there are no cut points of $C(a, b)$.

Remark. The reader familiar with the theory of cyclic chains in a Peano space will recognize this proof as standard. One of the results of Kuratowski and Whyburn is that if $a \sim b$ then $M(a, b) = C(a, b)$.[13] We shall see that this formula has to be somewhat modified in an arc-space. In fact, let us consider the example of §4.3. If z denotes the origin while a and b are points in the open square, then $C(a, b)$ is the open square and $M(a, b)$ is the whole arc-space. Hence $M(a, b) \neq C(a, b)$. The correct formula is that of §8.3. Notice that z is a stray frontier point of $C(a, b)$.

8.2. THEOREM. *If $a \sim b$ and $a \neq b$, then there are no cut points of $C^*(a, b)$.*

The proof is made by using §§8.1 and 7.11.

8.3. THEOREM. *If $a \sim b$ and $a \neq b$, then $C^*(a, b) = M(a, b)$.*

Proof. We shall first show that $M(a, b) \subset C^*(a, b)$. If $x \in' C^*(a, b)$, then $x \in S_x$ a component of $1 - C(a, b)$. Hence there is a frontier point z between $C(a, b)$ and S_x. (See §7.9.) If $z = x$ and is a stray frontier point, then $x \in C^*(a, b)$; hence if $z = x$, it is not a stray frontier point. This means that there is another, $z_1 \neq a$ say, between $C(a, b)$ and S_x which cuts the space between a and x. Hence a is not conjugate to x and so $x \in' M(a, b)$.

On the other hand, if $z \neq x$ and $z \neq a$ say, then a is not conjugate to x and $x \in' M(a, b)$. Hence $M(a, b) \subset C^*(a, b)$.

It is easy to see that $C^*(a, b) \subset M(a, b)$; for, if $x \in C^*(a, b)$, then by §8.2, $x \sim a$ and $x \sim b$, and this means that $x \in M(a, b)$.

COROLLARY. *A cyclic element M contains a frontier point of any component of $1 - M$.*

8.4. The relation between cyclic elements and connected sets is expressed in the following

THEOREM. *A non-degenerate connected set E is a true cyclic element if and only if it contains no cut points of itself and is saturated with respect to this property.*[14]

Proof. Suppose E is a true cyclic element $M(a, b)$ and a proper subset of a set A. By §6.2 no point of E cuts it so we have only to prove that some point of A cuts A. We may assume that A is connected, otherwise the fact is established. Take $x \in A$, $x \in' M(a, b)$. Now x is not conjugate to a or it is not

[13] See Kuratowski and Whyburn, loc. cit., p. 311.

[14] A set is said to be saturated with respect to a property if it has the property but is not a proper subset of any set having the property.

conjugate to b. Suppose that some point cuts the space between x and a. Since A is connected, this point must be a point of A and must cut A. (See §5.3.)

To prove the sufficiency, take $a \neq b$; $a, b \in E$. No point outside E can separate a from b. Moreover, for $p \in E$, $E - p$ is connected. Hence $a \sim b$. In fact, if $x \in E$, then $x \sim a$ and $x \sim b$. Thus $E \subset M(a, b)$. On the other hand, no point of $M(a, b)$ cuts $M(a, b)$ and so E cannot be a proper subset of $M(a, b)$ by hypothesis. Hence $E = M(a, b)$.

IX. Sets of type $\mathcal{H}$ and cyclic elements

9.1. It is important to know something of the structure of sets of type $\mathcal{H}$. In this connection we have the following

THEOREM. *If a connected set E has the property that any cyclic element having two points in E is contained entirely in E, then E is a set of type $\mathcal{H}$.*

Proof. Suppose that the theorem is false. Then there exist a pair of points $a, b \in E$ and an arc $\alpha(a, b)$ containing a point p which is not in E. Take any $\alpha_1(a, b) \subset E$. There is at least one such arc since E is connected. There is a first point of $\alpha(p, a)$ in common with $\alpha_1(b, a)$ and this point cannot be b, as $b \in' \alpha(p, a)$, nor can it be p, as $p \in' \alpha_1(b, a)$. Call it $a_1 \in E$. There is a first point on $\alpha(p, b)$ in common with $\alpha_1(a_1, b)$ and this point is distinct from a_1 and p. Call it $b_1 \in E$. Now $\alpha(a_1, b_1)$ and $\alpha_1(a_1, b_1)$ have only a_1 and b_1 in common. Hence $a_1 \sim b_1$ and p is conjugate to both. Therefore $p \in M(a_1, b_1)$, but $M(a_1, b_1) \subset E$ by hypothesis.

Remark. E need not be a set of type $\mathcal{H}$ which has the additional property that $E = E^*$.

9.2. THEOREM. *If H is of type $\mathcal{H}$, and a cyclic element $M(p, q)$ has two points in H, then $M(p, q) \subset H^*$.*

Proof. Suppose that $p, q \in H$. Now $M(p, q) = C^*(p, q)$. But $C(p, q)$ is contained in H by its very definition, and this implies, by §7.14, that $C^*(a, b) \subset H^*$.

X. Cyclic chains and cyclic elements

10.1. Following the customary notation, let $K(a, b)$ be the set of points cutting the space between a and b. Let $X(a, b) = a + K(a, b) + b + \sum M$, where the summation is taken over all cyclic elements M having two or more points in common with $a + K(a, b) + b$.

Consider any arc $\alpha(a, b)$. Now $a + K(a, b) + b$ is contained in $\alpha(a, b)$ and so it is ordered with respect to $\alpha(a, b)$ (see §§5.3 and 1.8). But more than this, it is easily seen that the ordering of $a + K(a, b) + b$ is independent of the arc $\alpha(a, b)$ in the sense that it is the same with respect to any two arcs $\alpha_i(a, b)$, for $i = 1, 2$. In this fashion $a + K(a, b) + b$ constitutes an ordered set which we shall designate by $\mathcal{K}(a, b)$. Two points of $\mathcal{K}(a, b)$ are called *consecutive* if

there is no point of $\mathcal{K}(a, b)$ between them. One might expect that by considering $\mathcal{K}(a, b)$ as an arc from a to b we again get an arc-space, but this is not the case as A4 certainly does not hold. However, the following statements can easily be proved.

10.2. If a true cyclic element M has two points in common with some $\alpha(a, b)$, it contains exactly two consecutive points of $\mathcal{K}(a, b)$.

10.3. Each pair of consecutive points in $\mathcal{K}(a, b)$ is contained in a true cyclic element.

10.4. If two true cyclic elements in $X(a, b)$ have a point in common, it is a point of $\mathcal{K}(a, b)$.

10.5. If $x \in \mathcal{K}(a, b)$, it is contained in at most two true cyclic elements of $X(a, b)$.

10.6. Any point on $\alpha(a, b)$ belongs to at least one and at most three cyclic elements in $X(a, b)$.

10.7. Any arc $\alpha(a, b)$ is contained in $X(a, b)$.

10.8. $X(a, b)$ is connected.

10.9. Given any arc $\alpha(a, b)$ and an arc $\alpha_1(p, q)$ with the property that it has a non-vacuous interior in $X(a, b)$ and that at most p and q are points of $\alpha(a, b)$, then $\alpha(p, q)$ is a subset of a single cyclic element in $X(a, b)$.

10.10. Of a somewhat more difficult character is the following

THEOREM. *If a cyclic element M has at least two points in common with $X(a, b)$, then it is in $X(a, b)$.*

Proof. Suppose that the theorem is false. Consider the case in which M has no points in common with $\mathcal{K}(a, b)$. Then it has exactly one point in common with each of two distinct true cyclic elements M_1 and M_2 in $X(a, b)$. If M_i has a_i and b_i in $\mathcal{K}(a, b)$, and $M \cdot M_i = p_i$ for $i = 1, 2$, we shall show that $p_2 \sim a_1, b_1$. Take any $x \in 1$ distinct from p_2 and a_1. If $x \neq p_1$, then as $p_2 \sim p_1$ and $p_1 \sim a_1$, it follows that p_2 and a_1 are in the same component of $1 - x$. If $x = p_1$, then $p_2 \sim a_2$; and as $p_1 \epsilon' \mathcal{K}(a, b)$, there is an arc omitting p_1 and joining a_1 to a_2. Therefore p_2 and a_1 are in the same component of $1 - x$. Hence $p_2 \sim a_1$. Similarly $p_2 \sim b_1$. Therefore $p_1, p_2 \in M_1$, by §6.1, and so $M = M_1$ by §6.2. Hence M does have points in $\mathcal{K}(a, b)$. The proof is similar in the other cases.

10.11. THEOREM. $X(a, b) \supset C(a, b)$.

Proof. $X(a, b)$ is a connected set having the property that if any cyclic element has two points in $X(a, b)$ it is contained entirely in $X(a, b)$. Hence by §9.1, $X(a, b)$ is a set of type $\mathcal{K}$ containing a and b. Therefore, by the definition of $C(a, b)$, $X(a, b) \supset C(a, b)$.

10.12. THEOREM. $X(a, b) \supset C^*(a, b)$.

Proof. Let z be a stray frontier point of $C(a, b)$. Then $z \epsilon' \alpha(a, b) \subset C(a, b)$. Let $p \epsilon \alpha(a, b)$ and consider any arc $\alpha_1(p, z)$. There is a first point of $\alpha_1(z, p)$ which is also on $\alpha(a, b)$. Call it $q \neq z$. Now $0 \neq \alpha_1(zq)^0 \subset C(a, b) \subset X(a, b)$. Hence by §10.9, $\alpha(z, q)$ belongs to a single cyclic element in $X(a, b)$.

10.13. THEOREM. $X(a, b) \subset C^*(a, b)$.

Proof. $C(a, b)$ is a set of type $\mathcal{H}$ containing a and b. Hence $C^*(a, b)$ contains any cyclic element having two points in common with $C(a, b)$ by §9.2. Hence $C^*(a, b)$ certainly contains $X(a, b)$.

10.14. As a consequence of §§10.12 and 10.13 we get the following

THEOREM.[15] $X(a, b) = C^*(a, b)$.

The importance of this theorem is that for any two points in 1 the set $X(a, b)$ when considered as a class of cyclic elements, is unique.

XI. The hyper-space Σ

11.1. If the class of cyclic elements of 1 is considered as a class of points, we call it a hyper-space Σ. Strictly speaking, the term hyper-space should not be applied until we have defined the meaning of the term arc for this collection of elements. We proceed to do this in such a fashion that the six axioms for an arc-space are satisfied, and, in addition, the space has the property that there is only one arc between any two points. Such an arc-space is called a *dendrite*.

A point in Σ will be denoted by a capital Latin letter enclosed in brackets; e.g., $[A]$. A will denote the set of points in 1 which constitute the cyclic element denoted by $[A]$ when it is considered as an element of Σ.

11.2. If $[A] \neq [B]$ are two points of Σ, then by an arc from A to B in 1 which is *proper at* A we mean an arc $\alpha(a, b)$ with the following properties: (1) $a \epsilon A$, $b \epsilon B$. (2) If A is non-degenerate, then at least two points of A are in $\alpha(a, b)$.

To obtain an arc proper at A, take $a \epsilon A$, $b \epsilon B$, $a \neq b$, and consider an arc $\alpha(a, b)$. If $a = A$, then $\alpha(a, b)$ is proper at A. If this is not the case, then two cases arise: (1) $b \epsilon A$. (2) $b \epsilon' A$. In (1), $\alpha(a, b)$ is proper by definition. In (2) a may be the only point of $\alpha(a, b)$ in A. If so it is a frontier point. Pick $a_1 \neq a$, $a_1 \epsilon A$. Now any arc $\alpha(a_1, b)$ contains a by §7.8 and so is proper.

By a proper arc from A to B we mean an arc proper at A and also at B. That the definition is non-vacuous may be shown as above. It follows that if A is non-degenerate, then any proper arc from A to B contains a frontier of A belonging to A, and so a cut point.

11.3. If $\alpha(a, b)$ is a proper arc from A to B, consider $[X(a, b)]$ as a collection of cyclic elements, not as points of 1. That is, $[X(a, b)]$ consists of: (1) the cyclic elements in $K(a, b)$, each of which is a point cutting the space between

[15] See Kuratowski and Whyburn, loc. cit., p. 316.

a and b. (2) The true cyclic elements which have two points in common with $\mathcal{K}(a, b)$. (3) The cyclic element a, in case $a = A$. (4) The cyclic element b, in case $b = B$.

Using the facts of Part X we see that the elements $[X(a, b)]$ have a definite order relative to the proper arc $\alpha(a, b)$. In fact, let us say $[P] < [Q]\ [\alpha(a, b)]$ if and only if the first point of $P \cdot \alpha(a, b)$ precedes the last point of $Q \cdot \alpha(a, b)$ along the arc $\alpha(a, b)$.

It is easy to see that the order above defined is independent of the proper arc. In other words $[X(a, b)]$ both in content and order depends only on A and B. It may further be shown that the first element is $[A]$ while the last is $[B]$.

We shall speak of the set of cyclic elements $[X(a, b)]$, defined and ordered as above, as an arc $\alpha([A], [B])$ from $[A]$ to $[B]$ in Σ.

It will be noticed that A6 is obviously satisfied, and in addition there is a unique arc between any two points of Σ.

In what follows, we shall see that for this definition of arc in Σ the six axioms of §2.4 are satisfied. In doing this the general plan will be to work with a proper arc $\alpha(a, b)$ in 1 and from its properties derive the analogous ones for $\alpha([A], [B])$. The two sections below will serve to illustrate this remark.

11.4. Suppose $[P]$ and $[Q]$ are elements of the arc $\alpha([A], [B])$ and $[P] < [Q]\ [\alpha([A], [B])]$. There is a proper arc $\alpha(a, b)$ from A to B in 1. If p is the first element of $P \cdot \alpha(a, b)$ and q is the last element of $Q \cdot \alpha(a, b)$ on $\alpha(a, b)$, then $\alpha(p, q)$ is a proper arc from P to Q. Now $[X(p, q)]$ is simply the totality of cyclic elements $[R]$ of $[X(a, b)]$ such that $[P] \leqq [R] \leqq [Q]\ [\alpha([A], [B])]$ and the ordering is consistent with the ordering of the cyclic elements in $[X(a, b)]$. Hence A2 is satisfied in Σ.

11.5. Next, suppose that the elements of an arc $\alpha([A], [B])$ in Σ are divided into two non-vacuous classes L' and U' such that $L' + U' = \alpha([A], [B])$, and $[P] \in L'$, $[Q] \in U'$ implies that $[P] < [Q]\ [\alpha([A], [B])]$.

Let $\alpha(a, b)$ be a proper arc from A to B and consider the points of 1 on $\alpha(a, b)$. Let $x \in L$ if the last cyclic element in $X(a, b)$ containing x is in L'. Let $x \in U$ otherwise. By §10.6 every point on $\alpha(a, b)$ must belong to at least one and can belong to at most three cyclic elements of $X(a, b)$. By A3 there is a last point in L or a first point in U, possibly both, along $\alpha(a, b)$.

If there is a last point in L, then there is clearly a last element in L'. If there is no last point in L, then there is a first point z in U. The last cyclic element of $X(a, b)$ along $\alpha(a, b)$ which contains z does not belong to L'; for, if it did, then z would be a point of L. On the other hand, no cyclic element of $X(a, b)$ whose last point along $\alpha(a, b)$ precedes z can belong to U'. Hence one of the cyclic elements containing z is the first of U'. (There may also be a last element of L'.) Hence A3 is satisfied in Σ.

11.6. The two preceding sections illustrate the general method of proof in showing that Σ is an arc-space. Σ satisfies the axioms, roughly speaking, because any proper arc in 1 satisfies the axioms.

It follows rather easily that $[A]/[B]$ if and only if one is a degenerate cyclic element contained in the other.

11.7. If we describe the operation of considering the hyper-space of an arc-space as reducing the arc-space modulo cyclic elements, then this operation can be applied *effectively* only once, in that after one application the hyper-space is a dendrite and remains a dendrite through further applications of the operation. An objection might be raised, in that while the character of the succeeding spaces does not change, since they are all dendrites, the number of points in a space may increase. For example, consider the Peano space consisting of two tangent circles as an arc-space. The hyper-space consists of three points of which one is a cut point and the other two are end points. But each end point is adjacent to a cut point and so we have two true cyclic elements in Σ. The hyper-space of Σ accordingly has five points, and so on.

The trouble is, of course, that we consider adjacent points as constituting a cyclic element. If we rule that a true cyclic element must consist of more than two points, then the theory goes over with hardly the change of a word and it is now possible to assert the final

THEOREM. *Reduction of an arc-space modulo cyclic elements leaves the hyper-space unchanged.*

THE OHIO STATE UNIVERSITY.

FUNCTIONS GENERATED BY DEVELOPING POWER SERIES IN CONTINUED FRACTIONS

By A. Markoff

This paper, although published long ago, remains unknown, evidently because of language difficulties. It contains important results concerning zeros of orthogonal polynomials, results susceptible of further developments. In view of the ever-growing interest in the field of orthogonal polynomials, it seems desirable to make this work of Markoff accessible to a larger group of readers by means of an English translation. The translator has adhered faithfully to the original, except for employing Perron's notation for continued fractions, the abbreviation $[\alpha_{i+j}]_0^{2n-2}$ for the determinant

$$\begin{vmatrix} \alpha_0 & \alpha_1 & \cdots & \alpha_{n-1} \\ \alpha_1 & \alpha_2 & \cdots & \alpha_n \\ \cdots & \cdots & \cdots & \cdots \\ \alpha_{n-1} & \alpha_n & \cdots & \alpha_{2n-2} \end{vmatrix},$$

and ${}_{i,j}[\alpha_{i+j}]_0^{2n-2}$ for the same determinant with the column α_i, α_{i+1}, $\cdots$ and the row α_j, α_{j+1}, $\cdots$ crossed out. In addition, he has numbered formulas and theorems in order to facilitate references.

Let

$$\frac{S_0}{x} + \frac{S_1}{x^2} + \frac{S_2}{x^3} + \cdots \tag{1}$$

be a power series arranged according to integral negative powers of the variable x.

By successive divisions it is possible to transform it, as is known, into the continued fraction

$$\frac{1|}{|q_1} - \frac{1|}{|q_2} - \cdots, \tag{2}$$

where q_1, q_2, $\cdots$ are polynomials in x.

These polynomials evidently depend also on the parameters S_0, S_1, S_2, $\cdots$.

We assign to the parameters

$$S_0, S_1, \cdots, S_{2m-2}, S_{2m-1} \tag{3}$$

real values for which none of the determinants

$$\Delta_1 = S_0, \qquad \Delta_2 = \begin{vmatrix} S_0 & S_1 \\ S_1 & S_2 \end{vmatrix}, \qquad \cdots, \qquad \Delta_m = [S_{p+q}]_0^{2m-2}$$

vanishes.

Received October 26, 1939. Originally published in the appendix to Memoirs (Zapiski) of the Imperial Academy of Sciences, vol. 74, no. 2(1894), pp. 1–30. Translated from the Russian by J. Shohat of the University of Pennsylvania.

Such being the case, $q_1, q_2, \cdots, q_k, \cdots, q_m$ are polynomials in x of the first degree. Accordingly, transforming, by the known rule, the fractions

$$\frac{1}{q_1}, \frac{1|}{|q_1} - \frac{1|}{|q_2}, \ldots, \frac{1|}{|q_1} - \frac{1|}{|q_2} - \cdots - \frac{1|}{|q_k}, \ldots, \frac{1|}{|q_1} - \frac{1|}{|q_2} - \cdots - \frac{1|}{|q_m}$$

into ordinary fractions

$$\frac{\varphi_1(x)}{\psi_1(x)}, \frac{\varphi_2(x)}{\psi_2(x)}, \cdots, \frac{\varphi_k(x)}{\psi_k(x)}, \cdots, \frac{\varphi_m(x)}{\psi_m(x)}, \tag{4}$$

we may take

$$\psi_k(x) = P_{0,k} + P_{1,k}x + P_{2,k}x^2 + \cdots + P_{k,k}x^k \tag{5}$$

and determine the ratios

$$\frac{P_{0,k}}{P_{k,k}}, \frac{P_{1,k}}{P_{k,k}}, \cdots, \frac{P_{k-1,k}}{P_{k,k}} \tag{6}$$

from the equations

$$S_iP_{0,k} + S_{i+1}P_{1,k} + \cdots + S_{i+k}P_{k,k} = 0 \qquad (i = 0, 1, \cdots, k-1)\cdot \tag{7}$$

The latter are equivalent to the condition that in the product

$$\psi_k(x)\left\{\frac{S_0}{x} + \frac{S_1}{x^2} + \cdots\right\}$$

the terms in $x^{-1}, x^{-2}, \cdots, x^{-k}$ vanish. Whence,

$$P_{i,k} = \frac{(-1)^{k-i}P_{k,k}}{\Delta_k}\,{}_{i,k}[S_{p+q}]_0^{2k-1}. \tag{8}$$

The function $\psi_k(x)$ once determined, it is easy to determine also

$$\varphi_k(x) = Q_{0,k} + Q_{1,k}x + \cdots + Q_{k-1,k}x^{k-1}, \tag{9}$$

which equals the integral part of the product

$$\psi_k(x)\left\{\frac{S_0}{x} + \frac{S_1}{x^2} + \cdots\right\}.$$

In fact, this last condition yields the following formulas:

$$\begin{cases} Q_{0,k} = S_0P_{1,k} + S_1P_{2,k} + \cdots + S_{k-1}P_{k,k}, \\ Q_{1,k} = \qquad S_0P_{2,k} + \cdots + S_{k-2}P_{k,k}, \\ \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\ Q_{k-1,k} = \qquad\qquad S_0P_{k,k}. \end{cases} \tag{10}$$

The coefficient $P_{k,k}$ remains as yet undetermined. This ambiguity, of course, disappears in the expression for the fraction $\varphi_k(x)/\psi_k(x)$. Moreover, if the

numerator and denominator of the fraction $\varphi_m(x)/\psi_m(x)$ are to be determined by the known formulas

$$\begin{aligned} \psi_1(x) &= q_1,\ \psi_2(x) = q_2\psi_1(x) - 1, \cdots, \psi_m(x) = q_m\psi_{m-1}(x) - \psi_{m-2}(x), \\ \varphi_1(x) &= 1,\ \varphi_2(x) = q_2, \qquad \cdots, \varphi_m(x) = q_m\varphi_{m-1}(x) - \varphi_{m-2}(x), \end{aligned} \tag{11}$$

the following equation must be satisfied:

$$\begin{aligned} (S_0P_{1,k} + S_1P_{2,k} + \cdots + S_{k-1}P_{k,k})P_{0,k-1} - P_{0,k}(S_0P_{1,k-1} + S_1P_{2,k-1} \\ + \cdots + S_{k-2}P_{k-1,k-1}) = 1, \end{aligned} \tag{12}$$

i.e.,

$$\varphi_k(0)\psi_{k-1}(0) - \varphi_{k-1}(0)\psi_k(0) = 1. \tag{13}$$

At this point, it is helpful to introduce the determinant

$$\Delta = \left| \begin{array}{c|c} & S_k \\ [S_{i+j}]_0^{2k-2} & S_{k+1} \\ & \cdots \\ & S_{2k-1} \\ \hline 0\ \ 0 \cdots 0 & 1 \end{array} \right| \tag{14}$$

and its minors $\Delta_{i,j}$ of the first order. $\Delta_{i,j}$ means $(-1)^{i+j}$ times the determinant obtained from Δ by crossing out the column S_i, S_{i+1}, S_{i+2}, $\cdots$, S_{i+k-1}, 0 (1, for $i = k$) and the row S_j, S_{j+1}, $\cdots$, S_{j+k} for $j = 0, 1, 2, \cdots, k - 1$; for $j = k$ we cross out the last row $0, 0, \cdots, 0, 1$.

With these notations, the ratios

$$\frac{P_{0,k}}{P_{k,k}}, \frac{P_{1,k}}{P_{k,k}}, \cdots, \frac{P_{k-1,k}}{P_{k,k}} \tag{15}$$

may be written in the form of fractions

$$\frac{\Delta_{0,k}}{\Delta}, \frac{\Delta_{1,k}}{\Delta}, \cdots, \frac{\Delta_{k-1,k}}{\Delta}, \tag{16}$$

and the ratios

$$\frac{P_{0,k-1}}{P_{k-1,k-1}}, \frac{P_{1,k-1}}{P_{k-1,k-1}}, \cdots, \frac{P_{k-2,k-1}}{P_{k-1,k-1}} \tag{17}$$

in the form of fractions

$$\frac{\Delta_{0,k-1}}{\Delta_{k-1}}, \frac{\Delta_{1,k-1}}{\Delta_{k-1}}, \cdots, \frac{\Delta_{k-2,k-1}}{\Delta_{k-1}}. \tag{18}$$

The equation (13), or its equivalent equation (12), we now rewrite as follows:

$$
\begin{aligned}
(19)\qquad &(S_0\Delta_{1,k} + S_1\Delta_{2,k} + \cdots + S_{k-2}\Delta_{k-1,k} + S_{k-1}\Delta_{k,k})\Delta_{0,k-1} \\
&\qquad - (S_0\Delta_{1,k-1} + S_1\Delta_{2,k-1} + \cdots + S_{k-2}\Delta_{k-1,k-1})\Delta_{0,k} = \frac{\Delta_k\Delta_{k-1}}{P_{k,k}P_{k-1,k-1}},
\end{aligned}
$$

or

$$
\begin{aligned}
(20)\qquad &S_0(\Delta_{1,k}\Delta_{0,k-1} - \Delta_{0,k}\Delta_{1,k-1}) + S_1(\Delta_{2,k}\Delta_{0,k-1} - \Delta_{0,k}\Delta_{2,k-1}) + \cdots \\
&\qquad + S_{k-2}(\Delta_{k-1,k}\Delta_{0,k-1} - \Delta_{0,k}\Delta_{k-1,k-1}) + S_{k-1}\Delta_{k,k}\Delta_{0,k-1} = \frac{\Delta_k\Delta_{k-1}}{P_{k,k}P_{k-1,k-1}}.
\end{aligned}
$$

On the other hand, as is known from the theory of determinants,

$$
(21)\qquad
\begin{aligned}
&\frac{\Delta_{1,k}\Delta_{0,k-1} - \Delta_{0,k}\Delta_{1,k-1}}{\Delta} = [S_{i+j}]_2^{2k-2}, \\
&\frac{\Delta_{2,k}\Delta_{0,k-1} - \Delta_{0,k}\Delta_{2,k-1}}{\Delta} = -\,{}_{2,k}[S_{i+j}]_1^{2k-1}, \\
&\frac{\Delta_{3,k}\Delta_{0,k-1} - \Delta_{0,k}\Delta_{3,k-1}}{\Delta} = {}_{3,k}[S_{i+j}]_1^{2k-1}, \\
&\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\
&\frac{\Delta_{k-1,k}\Delta_{0,k-1} - \Delta_{0,k}\Delta_{k-1,k-1}}{\Delta} = (-1)^{k-2}\,{}_{k-1,k}[S_{i+j}]_1^{2k-1},
\end{aligned}
$$

and

$$
\begin{aligned}
&S_0[S_{i+j}]_2^{2k-2} - S_1\,{}_{2,k}[S_{i+j}]_1^{2k-1} + \cdots + (-1)^{k-2}S_{k-2}\,{}_{k-2,k}[S_{i+j}]_1^{2k-1} \\
&\qquad\qquad + (-1)^{k-1}S_{k-1}\,[S_{i+j}]_1^{2k-3} = \Delta_k = \Delta.
\end{aligned}
$$

Therefore, condition (13) yields the following equations:

$$
(22)\qquad \Delta_k^2 = \frac{\Delta_k\Delta_{k-1}}{P_{k,k}P_{k-1,k-1}}, \qquad P_{k-1,k-1}P_{k,k} = \frac{\Delta_{k-1}}{\Delta_k}.
$$

Moreover, we must put $P_{1,1} = S_0^{-1} = \Delta_1^{-1}$ in order to get $\varphi_1(x) = 1$.

We thus find successively:

$$
P_{2,2} = \frac{\Delta_1^2}{\Delta_2},\ P_{3,3} = \frac{\Delta_2^2}{\Delta_1^2\Delta_3},\ P_{4,4} = \frac{\Delta_1^2\Delta_3^2}{\Delta_2^2\Delta_4},\ \cdots,
$$

and in general,

$$
(23)\qquad P_{2l,2l} = \frac{\Delta_1^2\Delta_3^2\cdots\Delta_{2l-1}^2}{\Delta_2^2\Delta_4^2\cdots\Delta_{2l-2}^2\Delta_{2l}},\ P_{2l+1,2l+1} = \frac{\Delta_2^2\Delta_4^2\cdots\Delta_{2l}^2}{\Delta_1^2\Delta_3^2\cdots\Delta_{2l-1}^2\Delta_{2l+1}}.
$$

The functions $\psi_k(x)$, $\varphi_k(x)$ are now completely determined. Finally, we derive $q_1, q_2, \cdots, q_k, \cdots, q_m$ as integral parts of the fractions

$$
\frac{\psi_1(x)}{1},\ \frac{\psi_2(x)}{\psi_1(x)},\ \cdots,\ \frac{\psi_k(x)}{\psi_{k-1}(x)},\ \cdots,\ \frac{\psi_m(x)}{\psi_{m-1}(x)}.
$$

We notice, without further discussion, that the coefficient of x in the expression for q_k is $P_{k,k}/P_{k-1,k-1}$.

Now turn to the roots of the equation

$$\psi_k(x) = 0, \tag{24}$$

which we denote by $x_{i,k}$, each particular root specified by the index i. We now represent the fraction $\varphi_k(x)/\psi_k(x)$ as[1]

$$\sum \frac{R_{i,k}}{x - x_{i,k}}, \tag{25}$$

where

$$R_{i,k} = \frac{\varphi_k(x_{i,k})}{\psi_k'(x_{i,k})}. \tag{26}$$

Expanding the sum (25) according to negative powers of x and keeping terms nvolving powers of $1/x$ less than $2k + 1$, we must get

$$\frac{S_0}{x} + \frac{S_1}{x^2} + \cdots + \frac{S_{2k-1}}{x^{2k}}.$$

We thus get the equations

$$\sum R_{i,k}x_{i,k}^l = S_l \qquad (l = 0, 1, \cdots, 2k - 1), \tag{27}$$

which serve to solve the problem about the increase or decrease of the zeros of (24) under some specified variation of the numbers

$$S_0, S_1, \cdots, S_{2k-1}. \tag{28}$$

First, we limit the parameters

$$S_0, S_1, \cdots, S_{2m-2} \tag{29}$$

by the inequalities

$$\Delta_1 > 0, \Delta_2 > 0, \cdots, \Delta_k > 0, \cdots, \Delta_m > 0. \tag{30}$$

With such limitations, all coefficients $P_{1,1}, P_{2,2}, \cdots, P_{k,k}, \cdots, P_{m,m}$ are positive, as seen from (23). Thus, the sequence

$$\psi_m(x), \psi_{m-1}(x), \cdots, \psi_k(x), \psi_{k-1}(x), \cdots, \psi_1(x), 1$$

has all properties of a Sturm sequence, and none of the equations

$$\psi_m(x) = 0, \psi_{m-1}(x) = 0, \cdots, \psi_k(x) = 0, \cdots, \psi_1(x) = 0 \tag{31}$$

has imaginary or multiple roots. Moreover, the numbers

$$R_{i,k} = \frac{\varphi_k(x_{i,k})}{\psi_k'(x_{i,k})} = \frac{1}{\psi_{k-1}(x_{i,k})\psi_k'(x_{i,k})} \tag{32}$$

are positive, since $\psi_{k-1}(x_{i,k})$ and $\psi_k'(x_{i,k})$ have the same sign.

[1] We exclude the case of multiple roots.

It is readily seen that all roots of each equation (31) are positive, if to the inequalities (30) we add

$$S_1 > 0,\ [S_{i+j}]_1^3 > 0,\ \cdots,\ [S_{i+j}]_1^{2m-1} > 0. \tag{33}$$

The subsequent discussion is confined, for brevity, to the zeros of the equation

$$\psi_m(x) = 0 \tag{34}$$

which we denote by x_i. We also replace $R_{i,m}$ by R_i. We proceed to solve the following problem: What are the variations of the zeros $\{x_i\}$ of the equation (34) under infinitely small variations of the parameters (3)? In other words, we wish to form the derivatives dx_i/dS_l and to determine their sign under the limitations stated above.

To this end, differentiate the equations (27). We get

$$\sum lR_i x_i^{l-1}\,dx_i + \sum x_i^l\,dR_i = dS_l \qquad (l = 0, 1, \cdots, 2m-1). \tag{35}$$

Form the determinant

$$\begin{vmatrix} 0 & \cdots & 0 & 1 & \cdots & 1 \\ 1 & \cdots & 1 & x_1 & \cdots & x_m \\ 2x_1 & \cdots & 2x_m & x_1^2 & \cdots & x_m^2 \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ lx_1^{l-1} & \cdots & lx_m^{l-1} & x_1^l & \cdots & x_m^l \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ (2m-1)x_1^{2m-2} & \cdots & (2m-1)x_m^{2m-2} & x_1^{2m-1} & \cdots & x_m^{2m-1} \end{vmatrix} = D \tag{36}$$

and its minors $D_{i,l}$ of first order ($i = 1, 2, \cdots, m$). Here $D_{i,l}$ is obtained from D by crossing out the column $0, 1, 2x_i, 3x_i^2, \cdots, (2m-1)x_i^{2m-2}$ and the row $lx_1^{l-1}, lx_2^{l-1}, \cdots, lx_m^{l-1}, x_1^l, x_2^l, \cdots, x_m^l$. With these notations, (35) gives

$$R_i \frac{dx_i}{dS_l} = (-1)^{i+l+1} \frac{D_{i,l}}{D}. \tag{37}$$

It remains to present $D_{i,l}$ and D in the simplest possible form. Here the determinant

$$T = \begin{vmatrix} 1 & 1 & \cdots & 1 & 1 & 1 & \cdots & 1 \\ t_1 & t_2 & \cdots & t_m & x_1 & x_2 & \cdots & x_m \\ t_1^2 & t_2^2 & \cdots & t_m^2 & x_1^2 & x_2^2 & \cdots & x_m^2 \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ t_1^{2m-1} & t_2^{2m-1} & \cdots & t_m^{2m-1} & x_1^{2m-1} & x_2^{2m-1} & \cdots & x_m^{2m-1} \end{vmatrix} \tag{38}$$

involving m arbitrary quantities $t_1, t_2, \cdots, t_m$ proves useful. In fact, a simple expression is known for T, namely,

$$T = (x_m - x_{m-1})(x_m - x_{m-2}) \cdots (x_m - t_1)(x_{m-1} - x_{m-2}) \cdots (t_2 - t_1), \tag{39}$$

and

$$D = \left(\frac{\partial^m T}{\partial t_1 \partial t_2 \cdots \partial t_m}\right)_{t_1 = x_1, \cdots, t_m = x_m}. \tag{40}$$

We thus find, without much difficulty,

$$(41)\qquad D = (-1)^{\frac{1}{2}m(m+1)}(x_m - x_{m-1})^4(x_m - x_{m-2})^4 \cdots (x_2 - x_1)^4.$$

Furthermore, if we arrange T according to powers of t_i and denote by $(-1)^{i+l+1}T_{i,l}$ the coefficient of t_i^l, then $D_{i,l}$ is

$$\left(\frac{\partial^{m-1} T_{i,l}}{\partial t_1 \partial t_2 \cdots \partial t_{i-1} \partial t_{i+1} \cdots \partial t_m}\right)_{t_1=x_1, \cdots, t_{i-1}=x_{i-1}, t_{i+1}=x_{i+1}, \cdots, t_m=x_m}$$

and $(-1)^{l+1}T_{i,l}/T_{i,2m-1}$ is the coefficient of t_i^l in the expansion in powers of t_i of the product

$$(42)\quad (t_i - t_1)(t_i - t_2) \cdots (t_i - t_{i-1})(t_i - t_{i+1}) \cdots (t_i - t_m)(t_i - x_1) \cdots (t_i - x_m).$$

Hence, $(-1)^{l+1}D_{i,l}$ equals the coefficient of t_i^l in

$$(43)\qquad \begin{aligned}&(t_i - x_1)(t_i - x_2) \cdots (t_i - x_{i-1})(t_i - x_{i+1}) \cdots \\ &\qquad\qquad (t_i - x_m)(t_i - x_1)(t_i - x_2) \cdots (t_i - x_m)\end{aligned}$$

multiplied by

$$(44)\qquad \left(\frac{\partial^{m-1} T_{i,2m-1}}{\partial t_1 \partial t_2 \cdots \partial t_{i-1} \partial t_{i+1} \cdots \partial t_m}\right)_{t_1=x_1, \cdots, t_{i-1}=x_{i-1}, t_{i+1}=x_{i+1}, \cdots, t_m=x_m}.$$

On the other hand, we have the following expression for $T_{i,2m-1}$:

$$(45)\qquad \begin{aligned}T_{i,2m-1} = (x_m - x_{m-1}) \cdots (x_m - t_{i+1})(x_m - t_{i-1}) \cdots (t_{i+1} - t_{i-1})\\ \cdots (t_3 - t_2)(t_2 - t_1).\end{aligned}$$

This, combined with the above results, yields

$$(46)\quad (-1)^{l+1}D_{i,l} = (-1)^{\frac{1}{2}m(m+1)+i}(x_m - x_{m-1})^4(x_m - x_{m-2})^4 \cdots (x_2 - x_1)^4 \cdot \frac{C_{i,l}}{\{\psi_m'(x_i)\}^2},$$

where $C_{i,l}$ is the coefficient of x^l in the expansion in powers of x of the product

$$(47)\quad P_{m,m}^2(x - x_1)^2 \cdots (x - x_{i-1})^2(x - x_i)(x - x_{i+1})^2 \cdots (x - x_m)^2 \equiv \frac{\psi_m^2(x)}{x - x_i}.$$

Hence,

$$(48)\qquad (-1)^{i+l+1}D_{i,l} = D \cdot \frac{C_{i,l}}{\{\psi_m'(x_i)\}^2},$$

$$(49)\qquad R_i \frac{dx_i}{dS_l} = (-1)^{i+l+1}\frac{D_{i,l}}{D} = \frac{C_{i,l}}{\{\psi_m'(x_i)\}^2}.$$

Assume now that the determinants (30), as well as (33), are all positive. Then all quantities

$$(50)\qquad x_i,\ R_i,\ (-1)^{l+1}C_{i,l}$$

are positive too, whence

$$(-1)^{l+1}\frac{dx_i}{dS_l} = (-1)^{l+1}\frac{C_{i,l}}{R_i\{\psi_m'(x_i)\}^2} > 0. \tag{51}$$

We have thus established the following

THEOREM I. *As long as the determinants*

$$\Delta_1 = S_0\,,\quad \Delta_2 = \begin{vmatrix} S_0 & S_1 \\ S_1 & S_2 \end{vmatrix}, \cdots,\quad \Delta_m = [S_{i+j}]_0^{2m-2}, \tag{52}$$

$$\Delta^{(1)} = S_1\,,\ \Delta^{(2)} = \begin{vmatrix} S_1 & S_2 \\ S_2 & S_3 \end{vmatrix}, \cdots,\ \Delta^{(m)} = [S_{i+j}]_1^{2m-1} \tag{53}$$

remain positive, the roots of the equation $\psi_m(x) = 0$ increase, if $S_1, S_3, \cdots, S_{2m-1}$ increase and $S_0, S_2, \cdots, S_{2m-2}$ decrease.

A similar conclusion holds for the determinants (52), (53).

THEOREM II. *If $S_1, S_3, S_5, \cdots, S_{2m-1}$ increase and $S_0, S_2, \cdots, S_{2m-2}$ decrease, then the determinants* (52) *decrease and those in* (53) *increase, as long as all these determinants remain positive.*

In order to prove this statement, observe that

$$(-1)^l\frac{d\Delta_m}{dS_l} \quad\text{and}\quad (-1)^{l+1}\frac{d\Delta^{(m)}}{dS_l} \tag{54}$$

may be represented as sums of determinants obtained from Δ_m and $\Delta^{(m)}$ respectively by crossing out one row and one column at whose intersection the parameter S_l is located. We proceed to show that all such determinants are positive as long as the determinants (52), (53) remain positive.

In fact, denoting by Ω any one of the former determinants, and introducing the above quantities R_i, x_i, we may put

$$\Omega = \begin{vmatrix} \sum R_i x_i^{\alpha_1+\beta_1} & \sum R_i x_i^{\alpha_1+\beta_2} & \cdots & \sum R_i x_i^{\alpha_1+\beta_{m-1}} \\ \sum R_i x_i^{\alpha_2+\beta_1} & \sum R_i x_i^{\alpha_2+\beta_2} & \cdots & \sum R_i x_i^{\alpha_2+\beta_{m-1}} \\ \cdots & \cdots & \cdots & \cdots \\ \sum R_i x_i^{\alpha_{m-1}+\beta_1} & \sum R_i x_i^{\alpha_{m-1}+\beta_2} & \cdots & \sum R_i x_i^{\alpha_{m-1}+\beta_{m-1}} \end{vmatrix},$$

where the integers $\alpha_1, \alpha_2, \cdots, \alpha_{m-1}, \beta_1, \beta_2, \cdots, \beta_{m-1}$ are subject to the inequalities $0 \leqq \alpha_1 < \alpha_2 < \cdots < \alpha_{m-1} \leqq m$; $0 \leqq \beta_1 < \beta_2 < \cdots < \beta_{m-1} \leqq m$; $\alpha_{m-1} - \alpha_1 \leqq m - 1$; $\beta_{m-1} - \beta_1 \leqq m - 1$.

By the theorem on multiplication of determinants, we may represent Ω as a sum of m determinants, each of which is a product of two determinants:

$$\Omega = R_1R_2\cdots R_{m-1}\begin{vmatrix} x_1^{\alpha_1} & \cdots & x_1^{\alpha_{m-1}} \\ \cdots & \cdots & \cdots \\ x_{m-1}^{\alpha_1} & \cdots & x_{m-1}^{\alpha_{m-1}} \end{vmatrix}\cdot\begin{vmatrix} x_1^{\beta_1} & \cdots & x_1^{\beta_{m-1}} \\ \cdots & \cdots & \cdots \\ x_{m-1}^{\beta_1} & \cdots & x_{m-1}^{\beta_{m-1}} \end{vmatrix} + \cdots$$

$$+ R_2R_3\cdots R_m\begin{vmatrix} x_2^{\alpha_1} & \cdots & x_2^{\alpha_{m-1}} \\ \cdots & \cdots & \cdots \\ x_m^{\alpha_1} & \cdots & x_m^{\alpha_{m-1}} \end{vmatrix}\cdot\begin{vmatrix} x_2^{\beta_1} & \cdots & x_2^{\beta_{m-1}} \\ \cdots & \cdots & \cdots \\ x_m^{\beta_1} & \cdots & x_m^{\beta_{m-1}} \end{vmatrix}.$$

On the other hand, if with arbitrary quantities $t_1, t_2, \cdots, t_{m-1}$ and with integers $\lambda_1, \lambda_2, \cdots, \lambda_{m-1}$, satisfying the inequalities

$$0 \leqq \lambda_1 < \lambda_2 < \cdots < \lambda_{m-1} \leqq m, \qquad \lambda_{m-1} - \lambda_1 \leqq m - 1,$$

we form the determinant

$$\begin{vmatrix} t_1^{\lambda_1} & t_1^{\lambda_2} & \cdots & t_1^{\lambda_{m-1}} \\ t_2^{\lambda_1} & t_2^{\lambda_2} & \cdots & t_2^{\lambda_{m-1}} \\ \cdots & \cdots & \cdots & \cdots \\ t_{m-1}^{\lambda_1} & t_{m-1}^{\lambda_2} & \cdots & t_{m-1}^{\lambda_{m-1}} \end{vmatrix}, \tag{55}$$

its ratio to the products $(t_1t_2 \cdots t_{m-1})^{\lambda_1}(t_{m-1} - t_{m-2})(t_{m-1} - t_{m-3}) \cdots (t_3 - t_2) \cdot (t_3 - t_1)(t_2 - t_1)$ certainly equals one of the following expressions:

$$1,\, t_1 + t_2 + \cdots + t_{m-1},\, t_1t_2 + t_1t_3 + \cdots + t_{m-2}t_{m-1},$$
$$\cdots,\, t_1t_2 \cdots t_{m-2} + \cdots + t_2t_3 \cdots t_{m-1}.$$

We may add that this ratio equals

$$\begin{aligned} &1, \text{ if } \lambda_2 - \lambda_1 = \lambda_3 - \lambda_2 = \cdots = \lambda_{m-1} - \lambda_{m-2} = 1; \\ &t_1 + t_2 + \cdots + t_{m-1}, \qquad \text{if } \lambda_{m-1} - \lambda_{m-2} = 2; \\ &t_1t_2 + \cdots + t_{m-2}t_{m-1}, \qquad \text{if } \lambda_{m-2} - \lambda_{m-3} = 2; \\ &\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\ &t_1t_2 \cdots t_{m-2} + \cdots + t_2t_3 \cdots t_{m-1}, \quad \text{if } \lambda_2 - \lambda_1 = 2. \end{aligned}$$

In all cases, this ratio is positive if $t_1, t_2, \cdots, t_{m-1}$ are positive. Therefore, for positive $t_1, t_2, \cdots, t_{m-1}$, the product of the two determinants of type (55), differing by the values of the exponents $\lambda_1, \lambda_2, \cdots, \lambda_{m-1}$ only, is certainly positive.

Thus, we may state that the determinants entering as summands in the above representation of Ω are all positive if

$$x_1 > 0, \cdots, x_m > 0, R_1 > 0, \cdots, R_m > 0. \tag{56}$$

This is certainly satisfied if

$$\Delta_1 > 0, \cdots, \Delta_m > 0, \Delta^{(1)} > 0, \cdots, \Delta^{(m)} > 0. \tag{57}$$

Thus, (57) implies the positiveness of all determinants Ω, hence also the positiveness of (54).

The above discussion concerns Δ_m and $\Delta^{(m)}$, but it is readily seen that the conditions (57) imply also

$$(-1)^l \frac{d\Delta_k}{dS_l} > 0, \qquad (-1)^{l+1} \frac{d\Delta^{(k)}}{dS_l} > 0 \quad (k = 1, 2, \cdots, m). \tag{58}$$

This is precisely the statement of Theorem II.

On the basis of the foregoing, we readily prove two remarkable theorems with which we close our paper. One concerns the determinants (52) and (53), the other the roots of the equation $\psi_m(x) = 0$.

Theorem III (On determinants). *Let the quantities* $S_0, S_1, \cdots, S_{2m-1}$ *assume two sets of values*: $S_i = a_i$, $S_i = b_i$ $(i = 0, 1, \cdots, 2m-1)$ *such that all determinants* Δ_k, $\Delta^{(k)}$ $(k = 1, 2, \cdots, m)$ *are positive, and let*

$$a_0 \geqq b_0, \; b_1 \geqq a_1, \cdots, a_{2m-2} \geqq b_{2m-2}, \; b_{2m-1} \geqq a_{2m-1}. \tag{59}$$

Then these determinants remain positive for all $S_0, S_1, \cdots, S_{2m-1}$ *such that*

$$a_0 \geqq S_0 \geqq b_0, \; b_1 \geqq S_1 \geqq a_1, \cdots, a_{2m-2} \geqq S_{2m-2} \geqq b_{2m-2}, \quad b_{2m-1} \geqq S_{2m-1} \geqq a_{2m-1}. \tag{60}$$

Moreover, under the same conditions

$$[a_{i+j}]_0^{2k-2} \geqq [S_{i+j}]_0^{2k-2} \geqq [b_{i+j}]_0^{2k-2} \tag{61}$$

$$[b_{i+j}]_1^{2k-1} \geqq [S_{i+j}]_1^{2k-1} \geqq [a_{i+j}]_1^{2k-1} \qquad (k = 1, 2, 3, \cdots, m). \tag{62}$$

Proof. For $m = 1$, the theorem is evident. Thus, if we wish to prove it for a certain value of m, we may assume it holds for smaller values of m. In other words, the theorem will be proved completely, if we establish the inequalities

$$[a_{i+j}]_0^{2k-1} \geqq \Delta_k \geqq [b_{i+j}]_0^{2k-2}, \tag{63}$$

$$[b_{i+j}]_1^{2k-1} \geqq \Delta^{(k)} \geqq [a_{i+j}]_1^{2k} \tag{64}$$

for $k = m$, assuming their validity for $k = 1, 2, \cdots, m-1$. Observe further that the determinants $\Delta_1, \cdots, \Delta_m$, $\Delta^{(1)}, \cdots, \Delta^{(m-1)}$ are independent of S_{2m-1}. Therefore, in proving (63) and (64) for $k = m$, we may assign an arbitrary value to S_{2m-1}, disregarding the inequalities $b_{2m-1} \geqq S_{2m-1} \geqq a_{2m-1}$. We now start increasing continuously $S_0, S_2, \cdots, S_{2m-2}$ and decreasing continuously $S_1, S_3, \cdots, S_{2m-3}$ from their initial values $S_i = b_i$ to the final values $S_i = a_i$ $(i = 0, 1, \cdots, 2m-2)$. In this manner, we evidently pass through any system of values $S_0, S_1, \cdots, S_{2m-2}$ satisfying (60). Assume for any such system of values the validity of the inequalities

$$\Delta_1 > 0, \cdots, \Delta_{m-1} > 0, \; \Delta^{(1)} > 0, \cdots, \Delta^{(m-2)} > 0, \tag{65}$$

$$\Delta^{(m-1)} \geqq [a_{i+j}]_1^{2m-3}. \tag{66}$$

Take now S_{2m-1} so large that

$$\Delta^{(m)} = \Delta^{(m-1)} S_{2m-1} + \begin{vmatrix} S_1 & S_2 & \cdots & S_{m-1} & S_m \\ S_2 & S_3 & \cdots & S_m & S_{m+1} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ S_{m-1} & S_m & \cdots & S_{2m-3} & S_{2m-2} \\ S_m & S_{m+1} & \cdots & S_{2m-2} & 0 \end{vmatrix}$$

remains positive. This is possible by (66). Under such conditions, Δ_m, by virtue of Theorem II, constantly increases if $S_0, S_2, \cdots, S_{2m-2}$ decrease and $S_1, S_3, \cdots, S_{2m-3}$ increase, for its initial value $[b_{i+j}]_0^{2m-2}$ is positive. Thus,

(63) is proved for $k = m$ if the S_i ($i = 0, 1, \cdots, 2m - 2$) all satisfy the inequalities (60) (omitting the last one). In the same way (65) and (66) are proved. We now turn to $\Delta^{(m)}$. Here we cannot assign an arbitrary value to S_{2m-1}. However, to the inequalities (65) we now add $\Delta_m > 0$. Under these conditions, if we decrease continuously $S_0, S_2, \cdots, S_{2m-2}$ and increase continuously $S_1, S_3, \cdots, S_{2m-1}$ from their initial values $S_i = a_i$ to the final values $S_i = b_i$ ($i = 0, 1, \cdots, 2m - 1$), $\Delta^{(m)}$, by virtue of the preceding, constantly increases, for its initial value $[a_{i+j}]_1^{2m-1}$ is positive. Thus, under our assumption, (64) is proved for $k = m$, for all values $S_0, S_1, \cdots, S_{2m-1}$ under consideration. This completely proves Theorem III.

THEOREM IV (On roots). *If the quantities* a_i, S_i, b_i ($i = 0, 1, \cdots, 2m - 1$) *satisfy all conditions of Theorem* III, *then the equations, of degree* m *in* x,

$$\begin{vmatrix} & [a_{i+j}]_0^{2m-2} & & \begin{matrix}1\\ x\\ \cdots\\ x^{m-1}\end{matrix} \\ a_m & a_{m+1} \cdots & a_{2m-1} & x^m \end{vmatrix} = 0, \qquad \begin{vmatrix} & [S_{i+j}]_0^{2m-2} & & \begin{matrix}1\\ x\\ \cdots\\ x^{m-1}\end{matrix} \\ S_m & S_{m+1} \cdots & S_{2m-1} & x^m \end{vmatrix} = 0,$$

$$\begin{vmatrix} & [b_{i+j}]_0^{2m-2} & & \begin{matrix}1\\ x\\ \cdots\\ x^{m-1}\end{matrix} \\ b_m & b_{m+1} \cdots & b_{2m-1} & x^m \end{vmatrix} = 0$$

have neither multiple, nor negative, nor complex roots. Moreover, the roots of the second equation are greater than the corresponding roots of the first one and smaller than the corresponding roots of the third one.[2]

By corresponding roots of our equations we mean the largest roots, the second largest roots, the third, and so on, finally, the smallest roots. This theorem is a simple corollary of Theorem III, for the equation

$$\begin{vmatrix} & [S_{i+j}]_0^{2m-2} & & \begin{matrix}1\\ x\\ \cdots\\ x^{m-1}\end{matrix} \\ S_m & S_{m+1} \cdots & S_{2m-1} & x^m \end{vmatrix} = 0$$

coincides with the equation $\psi_m(x) = 0$.

We remark in closing that an important particular case of Theorem IV (on roots) may be found in the recent paper of the academician P. L. Tchebycheff: *On the development in a continued fraction of a series arranged according to negative*

[2] [Here a misprint is corrected in the original. J. S.]

powers of the variable. The eminent scientist derives also a formula equivalent to ours

$$(-1)^{l+1}R_i\frac{dx_i}{dS_l} = (-1)^{l+1}\frac{C_{i,l}}{\{\psi_m'(x_i)\}^2}.$$

The method of the academician P. L. Tchebycheff is essentially different from ours.

Among other papers which have some points in common with the present one, I may mention one by Joachimstahl: *Bemerkungen über den Sturmschen Satz,* Journal für die reine und angewandte Mathematik, vol. 48, also one by Frobenius: *Ueber Relationen zwischen Näherungsbrüche von Potenzreihen,* ibid., vol. 90.

Among my own papers, I may mention two notes: *Sur les racines de certaines équations,* Mathematische Annalen, vol. 27, and my thesis: *On some applications of algebraic continued fractions.* My thesis, published in 1884, contains, among other results, the derivation of the formulas published without proof by the academician P. L. Tchebycheff in 1885 in a paper: *On the representation of limiting values of integrals by means of integral residues.*

THE DIRICHLET PROBLEM FOR THE DAMPED WAVE EQUATION

By D. G. Bourgin

In general the Dirichlet data problem for normal hyperbolic equations is overdetermined and no solution is possible since certain compatibility conditions possibly infinite in number must be satisfied. For special contours, however, this overdetermination does not arise. The main concern of this paper is the Dirichlet problem for the damped wave equation in one dimension. Besides a deeper insight into the distinction between the elliptic and hyperbolic type equations thus afforded, a striking interrelation of elementary number theory and differential equations is brought into evidence. The case of no damping has already been treated.[1] Although there are points of qualitative similarity between the present work and these earlier results, the inclusion of the damping term modifies the results and introduces many new features.

It is well known that the constant coefficient equation in two variables representing a damped wave may be reduced to the canonical form

$$L(y) \equiv y_{xx} - y_{tt} + K^2 y = 0. \tag{.1}$$

The data are given on the boundary Γ of the finite rectangle R,

$$0 \leqq x \leqq m\pi, \qquad 0 \leqq t \leqq T\pi. \tag{.2}$$

The term "rectangle", used in the sequel, unless otherwise qualified will always refer to the *closed* rectangle defined in (.2). We shall use the notation

$$\begin{gathered} \alpha = \frac{T}{m}, \qquad \beta = (Km)^2, \qquad \gamma = (KT)^2 = \alpha^2\beta, \\ \rho(l) = \left(1 - \frac{\beta}{l^2}\right)^{\frac{1}{2}}, \qquad \nu(l) = \left(1 + \frac{\gamma}{l^2}\right)^{\frac{1}{2}}, \end{gathered} \tag{.3}$$

where $l = 1, 2, \cdots$. Throughout the paper, we restrict ourselves to the case

$$\alpha < 1, \qquad \beta < 1. \tag{.4}$$

We denote by (A) and (B) the equations

$$\alpha = \frac{g}{l\rho(l)}, \tag{A}$$

$$\alpha = \frac{g\nu(l)}{l}. \tag{B}$$

Received November 23, 1939; presented to the American Mathematical Society, December 26, 1939.

[1] D. G. Bourgin and R. Duffin, *The Dirichlet problem for the vibrating string equation*, Bulletin of the American Mathematical Society, vol. 45(1939), p. 851. This paper is referred to below as B.D.

By a solution of (A) or (B) we shall mean a pair of *positive integers* g, l satisfying the equation in question.

The *multiplicity* of solutions of the Dirichlet problem depends directly on the number of solutions of (A) or (B). This is made plausible by the observation that when (A) is satisfied $2T\pi$ is an integral multiple of the period of a (pseudo) standing wave. For this wave the data on $t = T\pi$ provides no additional information not implied by the data on $t = 0$ and accordingly the amplitude and phase cannot both be determined. The *existence* of solutions of the Dirichlet problem is consequent on the degree of approximation of $\rho(l)$ or $\nu(l)$ by rational numbers. In the interests of compactness of exposition the pertinent results on the Diophantine problem associated with (A) and on the rational approximations to $\rho(l)$ and $\nu(l)$ are brought together in the first section. §2 treats the Dirichlet and cognate problems and makes effective use of the material in §1.

1. Preliminary lemmas.

LEMMA 0. *The existence of a solution of* (A) *implies and is implied by the existence of a solution of* (B).

The equations (A) and (B) are equivalent, respectively, to

$$\text{(A)}: \qquad (lT)^2 - (gm)^2 = (mTK)^2,$$

$$\text{(B)}: \qquad (gT)^2 - (lm)^2 = (mTK)^2.$$

Interchanging l and g carries (A) into (B) and marks the relation between solutions of these two equations. In view of Lemma 0 it is not necessary to make explicit mention of (B) in the sequel.

LEMMA 1. *If β is irrational, there is at most one solution of* (A).

In general no solution will exist. Plainly one solution may exist. Suppose two distinct solutions g_1, l_1 and g_2, l_2 are possible. Then

$$mg_1(l_1^2 - \beta)^{-\frac{1}{2}} = mg_2(l_2^2 - \beta)^{-\frac{1}{2}}, \tag{1}$$

whence

$$(g_1l_2)^2 - (g_2l_1)^2 = \beta(g_2^2 - g_1^2). \tag{1.1}$$

Since the left side of (1.1) is an integer and the right side is irrational, we are led to a contradiction.

LEMMA 2. *Let u, v, P be integers. Necessary and sufficient conditions that at least one positive integer solution x, y exist for*

$$(ux)^2 - (vy)^2 = P \tag{C}$$

are:

(a) *P contains the H.C.F. of u^2 and v^2,*

(b) *the H.C.F. of u^2 and P and the H.C.F. of v^2 and P are squares,*

(c) *$P/2^{2n}$ $(n \geqq 0)$ is odd,*

(d) *if* $P = abq^2r^2$, $u = \bar{u}q$, $v = \bar{v}r$ *where* $a > b \geqq 1$, *then, for some choice of* a *and* b, $a + b$ *is divisible by* $\bar{u}$ *and* $a - b$ *by* $\bar{v}$.

The necessity of conditions (a) and (b) is patent. On writing

$$(2) \qquad x = r\bar{x}, \qquad y = q\bar{y},$$

there results in place of equation (C)

$$(2.1) \qquad (\overline{ux} + \overline{vy})(\overline{ux} - \overline{vy}) = ab.$$

Hence

$$(2.2) \qquad \bar{x} = \frac{a+b}{2\bar{u}}, \qquad \bar{y} = \frac{a-b}{2\bar{v}}.$$

Thus (c) and (d) must be satisfied. Evidently as many solutions $\bar{x}$, $\bar{y}$ exist as there are factorizations of $P/(qr)^2$ satisfying (2.2). Here and in the two lemmas following, the sufficiency of the conditions imposed is obvious since the steps in the necessity argument may be retraced.

LEMMA 3. *If* $\alpha = r/s < 1$ *where* r *and* s *are relatively prime, then necessary and sufficient conditions for the existence of a solution of* (A) *are that* (a) $\beta = P/Q < 1$, *where* P *and* Q *are relatively prime integers,* (b) $Q = u^2$, (c) $r = Gu$, (d) 2^{2n} *is the highest power of* 2 *in* P, (e) *if* $s = \bar{s}q$ *then for some integers* $a > b \geqq 1$, $P = abq^2$ *and* $\frac{1}{2}(a+b)/v$, $\frac{1}{2}(a-b)/\bar{s}$ *are integers. There is at most a finite number of solutions of* (A).

Evidently if a solution of (A) exists, then

$$(3) \qquad \left(\frac{r}{gs}\right)^2 = \frac{1}{l^2 - \beta}.$$

Condition (a) is therefore obvious. Moreover, if we write (3) in the form

$$(3.1) \qquad \left(\frac{r}{gs}\right)^2 = \frac{Q}{Ql^2 - P},$$

then, evidently, the numerator and denominator of the fraction on the right of (3.1) are each relatively prime perfect squares. Hence (b) must be satisfied. We write

$$(3.2) \qquad g = Gv, \qquad r = Gu,$$

where $G \geqq 1$. Hence

$$(3.3) \qquad (sv)^2 = (ul)^2 - P.$$

Here s and u are relatively prime by the second of the relations in (3.2). Also P and u are relatively prime. Now (3.3) is of the form (C) and our previous lemma applies. Thus (d) and (e) are justified.

LEMMA 4. *Let* $\alpha = (r/s)^{\frac{1}{2}} < 1$, *where the relatively prime integers* r *and* s *are not both squares. Necessary and sufficient conditions for a solution of* (A) *are*

(a) $\beta = P/Q < 1$, *where Q is prime to P and s*, (b) $r = G^2Q$, (c) *Q and s are not both squares*, (d) *if* p^{2n+1} *and* p^m *are the highest powers of the prime p in P and s respectively, then* $2n + 1 \geqq m$, (e) *if* $P = ed^2D^2\bar{P}$, $s = ed^2\bar{s}$, *where* $\bar{s}$ *is prime to D and* $\bar{P}$, *the equation* $eQu^2 - (\bar{s}\bar{v})^2 = \bar{P}$ *has an integer solution* $(u, \bar{v})$. *If one solution of* (A) *exists, then an infinite number exists.*

Condition (a) is plainly required. Equation (A) becomes

$$\frac{r}{s} = \frac{g^2Q}{Ql^2 - P}. \tag{4}$$

The highest common factor of g^2 and $Ql^2 - P$ must be a perfect square and we denote it by v^2. Hence it is necessary that an integer G, $G \geqq 1$, exist such that

$$g = Gv, \tag{4.1}$$

$$r = G^2Q, \tag{4.11}$$

$$Ql^2 - v^2s = P. \tag{4.12}$$

Now Q is prime to s by (4.11). Q is also prime to v, for otherwise, according to (4.12), P would not be prime to Q. Q and s cannot both be squares, for this would imply by (4.11) that r and s are both squares. The substitution

$$l = edDu, \qquad v = D\bar{v} \tag{4.3}$$

reduces (4.12) to the form

$$eQu^2 - (\bar{s}\bar{v})^2 = \bar{P}. \tag{4.4}$$

Clearly a necessary condition to be satisfied is that $\bar{s}$ and e be relatively prime and this is tantamount to condition (d). Moreover, any solution of (4.4) satisfies the condition that Q and $\bar{v}$ are relatively prime, for otherwise P and Q would have a common factor. It is a classical result, following on the Pell equation, that one solution of (4.4) implies the existence of an infinity of others.[2]

LEMMA 5. *If a solution of* (A) *exists, then the rationality of* α^2 *implies and is implied by the rationality of* β.

LEMMA 6. *If* $\alpha = r/s$ *where r and s are positive integers, then there is an integer L and a constant C such that for each l,* $l > L$, *there exists an associated integer* N_l *for which* $\frac{1}{2} > | r\rho(l)/s - N_l/l | \geqq C/l^2$. *For* $l \not\equiv 0 \pmod{s}$, C/l^2 *may be replaced by* C'/l.

Evidently

$$\frac{\beta}{l^2} \geqq 1 - \rho(l) \geqq \frac{\beta}{3l^2}. \tag{6}$$

Thus

$$\frac{\beta}{l^2}\frac{r}{s} \geqq \frac{r}{s} - \frac{r\rho(l)}{s} \geqq \frac{r\beta}{3sl^2}. \tag{6.1}$$

[2] E. Cahen, *Théorie des Nombres*, vol. 2, Paris, 1924, Chapter 13.

The nearest integer N_l to $lr\rho(l)/s$ is either (a) identical with lr/s or (b) differs from lr/s by at least s^{-1}. For case (a) the assertion of the lemma is contained in (6.1). In case (b)

$$\begin{aligned}\frac{1}{2} &\geqq \left| N_l - \frac{rl\rho(l)}{s} \right| \geqq \left| N_l - \frac{rl}{s} \right| - \left| \frac{rl}{s} - \frac{rl}{s}\rho(l) \right| \\ &\geqq \frac{1}{s} - \frac{r\beta}{sl} \geqq \frac{1}{2s}, \qquad l \geqq L = 2r\beta.\end{aligned} \tag{6.2}$$

LEMMA 7. *If $\alpha = r/s$, where r and s are relatively prime integers, there is an integer L and a constant C such that for each $l \geqq L$, there exists an associated integer N_l with the property*

$$\frac{1}{2} \geqq \left| \frac{\nu(l)}{\alpha} - \frac{N_l}{l} \right| \geqq \frac{C}{l^2}.$$

The demonstration is similar to that of Lemma 6. It is worth while to remark that, both here and in Lemma 6, β may be *any* real number inferior to 1.

LEMMA 8. *If β and α are algebraic numbers of degrees R and S respectively,*[3] *then there is an integer L and a constant C such that for each $l \geqq L$ there exists an associated integer N_l with the property*[4]

$$\frac{1}{2} \geqq \left| \alpha\rho(l) - \frac{N_l}{l} \right| \geqq \frac{C}{l^{2RS}}.$$

Suppose $\{\alpha_i\} = \alpha_1, \cdots, \alpha_S$ and $\{\beta_i\} = \beta_1, \cdots, \beta_R$, where $\alpha = \alpha_1$ and $\beta = \beta_1$ are the roots of irreducible[5] algebraic equations, of degrees S and R respectively, with integral coefficients. The equations satisfied by β_i and $l^2 - \beta_i$ are equivalent under the transformation $x' = l^2 - x$. That is to say, $l^2 - \beta_i$ satisfies an R-th degree equation with integer coefficients. Let $(r_\mu(l))^2$ ($\mu = 1, \cdots, RS$) stand for the products $\alpha_j^2(1 - \beta_i/l^2)$ ($i = 1, \cdots, R; j = 1, \cdots, S$), ordered by a suitable convention. The sums over μ of the 2σ-th powers of $r_\mu(l)$ are the products of the corresponding sums formed for $\{\alpha_i\}$ and $\{(1 - \beta_i/l^2)^{\frac{1}{2}}l\}$. Thus the elementary symmetric functions of degree 2σ formed for $\{r_\mu(l)\}$ are rational numbers represented as polynomials of maximum degree 2σ in l with rational coefficients. It is easy to show that the expressions $r_\mu(l)$ satisfy

$$F_l(x) \equiv A_0(lx)^{2RS} - A_2(l)(lx)^{2(RS-1)} + \cdots + A_{2RS}(l) = 0, \tag{8}$$

where $A_{2k}(l)$ is a polynomial of maximum degree $2k$ in l, with integer coefficients.

[3] For $S = 1$ Lemma 8 is much cruder than Lemma 6.

[4] The exponent $2RS$ in the conclusion may in some cases be replaced by a smaller number. This is possible if, for instance, α is the $2S$-th root of a rational number. More generally the choice $A = \alpha^2$, $B = \alpha^2\beta$ in Lemma 9 yields the exponent $2R'S$, where R' is the degree of $\alpha^2\beta$. If $\alpha^2 = A_1A_2$, $\beta = B_1/A_2$ where the degrees of A_1, A_2, B_1 are M_1, M_2, N_1, then it is easy to show that the exponent $2RS = 2M_2M_1N_1$ may be used.

[5] In this paper reducibility is considered solely with respect to the field of rational numbers.

Moreover,

$$(8.1)\qquad | A_{2\sigma}(l) | \leqq Cl^{2\sigma}.$$

The irrationality of the α's and β's for R, $S > 1$ carries with it the irrationality of $\{r_\mu(l)\}$ for $l \geqq l_1$. Since $A_{2\sigma}(l)$ is an integer, it follows that for all integers N

$$(8.2)\qquad \left| F_l \frac{N}{l} \right| \geqq 1, \qquad l \geqq l_1.$$

In view of (8.1)

$$(8.3)\qquad \begin{aligned} \left| \frac{d}{dx} F_l(x) \right| &\leqq \max_{0 \leqq \sigma < RS} | A_{2\sigma}(l) 2\sigma (2\alpha_1)^{2(RS-\sigma)} | RS, \qquad 0 < x < 2\alpha_1, \\ &\leqq Ml^{2RS}. \end{aligned}$$

The value of N which minimizes

$$\left| \frac{N}{l} - \alpha_1 \left(1 - \frac{\beta_1}{l^2}\right)^{\frac{1}{2}} \right|$$

is denoted by N_l. Plainly

$$(8.4)\qquad 0 < \frac{N_l}{l} < 2\alpha_1 \quad \text{for} \quad l > L > l_1,$$

$$(8.5)\qquad \left| \frac{N_l}{l} - \alpha_1 \left(1 - \frac{\beta_1}{l^2}\right)^{\frac{1}{2}} \right| < \frac{1}{2}.$$

Manifestly

$$(8.6)\qquad 0 < \alpha\rho(l) < \alpha.$$

Taylor's formula yields

$$(8.7)\qquad 1 \leqq \left| F_l\left(\frac{N_l}{l}\right) \right| \leqq \left| \frac{N_l}{l} - \alpha\rho(l) \right| \left| \frac{dF_l(x)}{dx} \right|$$

for x between N_l/l and $\alpha\rho(l)$ and $l > L$. Hence, in view of (8.3), (8.4), (8.5), (8.6) and (8.7),

$$(8.8)\qquad \frac{1}{2} \geqq \left| \frac{N_l}{l} - \alpha\rho(l) \right| \geqq \frac{1}{Ml^{2RS}}, \qquad l > L.$$

This is the assertion of the lemma.[6]

LEMMA 9. *If A and B are algebraic numbers of degrees S and R′ respectively, then there exist fixed numbers M and L such that, for all N,*

$$\left| \left(A + \frac{\beta}{l^2}\right)^{\frac{1}{2}} - \frac{N}{l} \right| > \frac{1}{Ml^{2R'S}}, \qquad l > L.$$

[6] A stronger result may be obtained by generalizing the Thue-Siegel inequality. C. Siegel, *Über Näherungswerte algebraischer Zahlen*, Mathematische Annalen, vol. 84(1921), pp. 80–99.

The demonstration may be patterned on that for Lemma 8.

LEMMA 10. *If α and β are algebraic numbers of degrees S and R respectively, then there are an integer L and a constant C such that $\frac{1}{2} \geqq | \nu(l)/\alpha - N/l | > C/l^{2RS}$ for each $l > L$ and all N.*

We start with the identity

$$\left|\frac{\nu(l)}{\alpha} - \frac{N}{l}\right| \equiv \left|\left(\frac{1}{\alpha^2} + \frac{\beta}{l^2}\right)^{\frac{1}{2}} - \frac{N}{l}\right|. \tag{10}$$

Clearly α^2 and α^{-2} are algebraic numbers of the same degree. Furthermore if α is of degree S, then α^2 is at most of degree S and may be of degree $\frac{1}{2}S$. Hence, on identifying α^{-2} and β with A and B of Lemma 9, the conclusion desired follows at once (see footnote 4).

LEMMA 11. *Let α be an algebraic number of degree S and let* (A) *have a solution. Then β and α^2 are algebraic numbers of the same degree, either S or $\frac{1}{2}S$. For each degree β satisfies some one of a two-parameter family of algebraic equations (with rational coefficients) of that degree determined by α.*

Let $F(x)$ be the irreducible polynomial of degree S with rational coefficients which is satisfied by $x = 1/\alpha$. Then

$$F_{gl}(x) \equiv F\left(\frac{xl}{g}\right) = F_{gl}^{(1)}(x^2) + xF_{gl}^{(2)}(x^2) = 0 \tag{11}$$

is satisfied by $x = g/l\alpha$. A straightforward argument shows that either

$$(F_{gl}^{(1)}(1 - z))^2 - (1 - z)(F_{gl}(1 - z))^2 = 0 \tag{11.1}$$

or

$$F_{g,l}^{(1)}(1 - z) = 0 \tag{11.2}$$

is satisfied by β/l^2 depending on the degree of α^2. It is evident from the form of (A) that α^2 and β may be interchanged in the hypotheses and conclusion.

A trivial corollary of the lemma is that either both or neither of α and β are transcendental numbers if (A) has a solution.

The conditions in the lemma, though shown to be necessary, are *not sufficient* for the existence of a solution of (A). The existence of l, g for which either (11.3) or (11.5) is satisfied by β/l^2 implies that *some one* of the S roots of (11) or of $F_{g,l}(-x) = 0$ is identical with $\rho(l)$. However, unless further restrictions are introduced to guarantee that the root just mentioned is actually $g/l\alpha$, we cannot assert that (A) is satisfied.

LEMMA 12. *If for some constant C and each $l \geqq L$ there exists an integer N_l such that*

$$\frac{1}{2} \geqq \left|\alpha\rho(l) - \frac{N_l}{l}\right| > \frac{C}{l^{w+1}},$$

then

$$|\sin \alpha\pi l\rho(l)| > \frac{D}{l^w}, \qquad l \geqq L.$$

If for some choice C and all $l \geqq L$ there exists an integer N_l such that

$$\frac{1}{2} \geqq \left| \frac{\nu(l)}{\alpha} - \frac{N_l}{l} \right| > \frac{C'}{l^{w+1}},$$

then

$$\left| \sin \frac{\pi l \nu(l)}{\alpha} \right| > \frac{D'}{l^w}, \qquad l \geqq L.$$

The proof is straightforward.

LEMMA 13. *If α and β are algebraic numbers of degrees S and R respectively* (see footnote 4), *then*

$$|\sin \alpha\pi l\rho(l)| > \frac{D}{l^{2RS-1}}, \qquad l \geqq L,$$

$$\left| \sin \frac{\pi l \nu(l)}{\alpha} \right| > \frac{D}{l^{2RS-1}}, \qquad l \geqq L.$$

This is a direct consequence of Lemmas 8, 9 and 12.

2.[7] In order to avoid lengthy circumlocution we introduce a convenient terminology. By an *admissible function* $y(x, t)$ we shall mean that (a) $y(x, t)$ is a solution of (.1) in the interior of the rectangle, (b) $y(x, t)$ is of class[8] C^1 in the rectangle, (c) y_{xx} and y_{tt} are Lebesgue summable over the rectangle, and (d) $y_x(x, t)$ is absolutely continuous in x for almost all t, $0 \leqq t \leqq T\pi$, and $y_t(x, t)$ is absolutely continuous in t for almost all x, $0 \leqq x \leqq m\pi$. Suppose $y_1, y_2, \cdots$ constitute a finite or denumerably infinite set of linearly independent admissible functions. The totality of all linear combinations with constant coefficients of a finite number of these functions will be referred to as the *linear manifold* determined by $\{y_i\}$ and will be denoted by the symbol $M\{y_i\}$. The number of functions in $\{y_i\}$ is the dimension of the manifold. By saying that $\{y_i\}$ is closed in the normed vector space Q we shall mean that $M\{y_i\}$ is dense in Q.

LEMMA 14. *If $y(x, t)$ is continuous in both real arguments in the rectangle, and if $F(p, u)$ is defined as*

$$F(p, u) = \int_0^{T\pi} \int_0^{m\pi} y(x, t) e^{i(xp+tu)}\, dx\, dt,$$

then $F(p, u)$ is an entire function in each of the complex variables p and u.

[7] Equations in §2 are given numbers above 20.

[8] The class C^n comprises functions continuous together with their derivatives through those of order n in both variables.

THEOREM 1. *A necessary and sufficient condition that the only admissible function vanishing on the boundary of R be $y(x, t) \equiv 0$ is that* (A) *have no solution.*

We take up the sufficiency demonstration first. The Green's theorem is evidently valid in the form

$$\begin{aligned}(20.1)\quad &\int_0^{T\pi}\int_0^{m\pi} [e^{i(px+ut)}L(y) - yL(e^{i(px+ut)})]\,dx\,dt \\ &= \int_0^{T\pi} e^{i(px+ut)}(y_x - ipy)\Big|_{x=0}^{m\pi} dt - \int_0^{m\pi} e^{i(px+ut)}(y_t - ivy)\Big|_{t=0}^{T\pi} dx.\end{aligned}$$

Since $y|_\Gamma = 0$ there results

$$(20.11)\quad (p^2 - u^2 - K^2)F(p, u) = \int_0^{T\pi} e^{i(px+ut)} y_x\Big|_{x=0}^{m\pi} dt - \int_0^{m\pi} e^{i(px+ut)} y_t\Big|_0^{T\pi} dx.$$

Since $F(p, u)$ is finite by Lemma 14, it follows that the right side of (20.11) vanishes for $u = \pm u_1$ where $u_1 = (p^2 - K^2)^{\frac{1}{2}}$. That is to say,

$$(20.2)\quad \int_0^{T\pi} e^{i(px\pm u_1t)} y_x\Big|_{x=0}^{m\pi} dt - \int_0^{m\pi} e^{i(px\pm u_1t)} y_t\Big|_0^{T\pi} dx = 0.$$

In consequence of (20.2)

$$(20.21)\quad \sin m\pi p \int_0^{T\pi} \sin tu_1\, y_x(m\pi, t)\, dt = \sin T\pi u_1 \int_0^{m\pi} \sin px\, y_t(x, T\pi)\, dx.$$

With the special values

$$(20.3)\quad p = \frac{l}{m} \qquad (l = 1, 2, \cdots)$$

we see from (20.21) that

$$(20.22)\quad \left[\int_0^{m\pi} \sin\frac{lx}{m}\, y_t(x, T\pi)\, dx\right] \sin \alpha l\rho(l)\pi = 0.$$

Now

$$(20.4)\quad \sin \alpha l\pi\rho(l) \neq 0 \qquad (l = 1, 2, \cdots),$$

for (A) has no solutions. Hence

$$(20.5)\quad \int_0^{m\pi} \sin\frac{lx}{m}\, y_t(x, T\pi)\, dx = 0 \qquad (l = 1, 2, \cdots).$$

Since, by the definition of admissible functions, $y_t(x, T\pi)$ is certainly of class C, it can be expanded over the range $0 \leqq x \leqq m\pi$ in a Fourier sine series, summable $(C, 1)$ to it at every point in the range. Thus, according to (20.5),

$$(20.6)\quad y_t(x, T\pi) = 0, \qquad 0 \leqq x \leqq m\pi.$$

In view of (20.21) and Lemma 0 the special choice

$$u_1 = \frac{l}{T} \qquad (l = 1, 2, \cdots) \tag{20.31}$$

yields

$$y_x(m\pi, t) = 0, \qquad 0 \leqq t \leqq T\pi. \tag{20.61}$$

On combining (20.2), (20.6) and (20.61) we establish

$$\int_0^{T\pi} e^{\pm iu_1 t} y_x(0, t)\, dt = \int_0^{m\pi} e^{ipx} y_t(x, 0)\, dx. \tag{20.7}$$

The special choices of p and u_1 defined in (20.3) and (20.31) then lead to

$$y_x(0, t) \equiv 0, \qquad 0 \leqq t \leqq T\pi, \tag{20.62}$$

$$y_t(x, 0) \equiv 0, \qquad 0 \leqq x \leqq m\pi. \tag{20.63}$$

Manifestly then

$$(p^2 - u^2 - K^2)F(p, u) = 0. \tag{20.8}$$

Accordingly $F(p, u)$ vanishes except possibly for the planes $u = \pm u_1$. Since $F(p, u)$ is certainly continuous in both complex variables, $F(p, u)$ must vanish for $u = \pm u_1$ also. That is to say, $F(p, u)$ vanishes identically. On recalling its definition as the Fourier transform of $y(x, t)$ it follows that

$$y(x, t) \equiv 0. \tag{20.9}$$

The necessity demonstration is immediate. If (A) has the solution $\bar{g}$, l, then $y(x, t)$, where

$$y(x, t) = \sin \frac{lx}{m} \sin \frac{tl\rho(l)}{m},$$

is an admissible non-identically vanishing solution satisfying the boundary conditions.

In view of the importance of Theorem 1 and in order to present a useful alternative approach to the questions of interest in this paper we give another demonstration of the theorem later in the paper. The theorem can be given more precise form by introducing the conditions in Lemmas 1, 3, 4, 11 or 12.

THEOREM 2. *If* $Y(\Gamma) \in C^1$ *and vanishes at the corners, and if* (A) *has no solution, then there is at most one admissible solution,* $y(x, t)$, *taking on the boundary values* $Y(\Gamma)$.

We write

$$y^* = y_1 - y_2, \tag{21}$$

where y_1 and y_2 are admissible solutions. Then y^* satisfies the conditions of Theorem 1 and accordingly vanishes identically.

Our previous lemmas, for example, Lemmas 1, 3, 4 and 11, interrelate the numerical nature of α and β and the number of solutions of (A). We are then led to a variety of interesting theorems connected with the multiplicity of admissible solutions of our Dirichlet problem.

THEOREM 3. *If β is irrational, then the Dirichlet problem with zero data on the boundary of the rectangle has either the unique admissible solution $y \equiv 0$ or there is a one-dimensional linear manifold of solutions.*

THEOREM 4. *If α and β are rational, then the Dirichlet problem, for 0 data on the boundary of the rectangle, has either the unique admissible solution $y \equiv 0$ or the admissible solutions determine a finite linear manifold.*

THEOREM 5. *If α^2 and β are rational, then the Dirichlet problem with 0 data on the boundary of the rectangle has either the unique solution $y = 0$ or there is an infinite-dimensional manifold of solutions.*

Before proving these theorems we remark that we have not made use of the full precision of Lemmas 1, 3 and 4. Plainly, the preceding three theorems can be made sharper by adding those hypotheses in the lemmas quoted, which distinguish between the nonexistence of solutions of (A) and the existence of a finite or infinite number of such solutions. We observe further that, if certain independent auxiliary conditions are imposed on $y(x, t)$, equal in number to the dimensionality of the linear manifolds of admissible solutions, the preceding theorems may be restated as uniqueness theorems. For instance, Theorem 4 may be put in the form:

THEOREM 4′. *Let α and β be rational numbers. Let $\{l_i, g_i\}$ $(i = 1, \cdots, N)$ be the solutions of (A) determined by the character of α and β as specified in the hypotheses of Lemma 3. Suppose the admissible functions satisfy*

$$a_i = \int_0^{T\pi} \int_0^{m\pi} y(x, t) \sin \frac{l_i x}{m} \sin \frac{g_i t}{T}\, dx\, dt \qquad (i = 1, \cdots, N),$$

where the constants $\{a_i\}$ are arbitrary. Then the Dirichlet problem, with $y \mid_\Gamma = 0$, has a unique solution. If $a_i \underset{i}{\equiv} 0$ this solution is $y(x, t) \equiv 0$.

We restrict ourselves to the proof of Theorem 4 since the argument for the other theorems differs in no essential way. If $\{g_i, l_i\}$ $(i = 1, \cdots, N)$ are solutions of (A), then from (20.22) we get as the most general value of $y_t(x, T\pi)$

$$y_t(x, T\pi) = \sum_{i=1}^{N} c_{l_i} \sin \frac{l_i x}{m}. \tag{22}$$

Let us suppose then that $\bar{y}(x, t)$ is an admissible solution of our Dirichlet problem for $\bar{y} \mid_\Gamma = 0$ and furthermore that

$$\bar{y}_t(x, T\pi) = 0. \tag{22.1}$$

It is then clear from the analysis in Theorem 1, subsequent to (20.6), that

$\bar{y}(x, t) \equiv 0$ is the unique admissible solution. Consider now the admissible solution

$$(22.2) \qquad y'(x, t) = \sum_{i=1}^{N} \frac{mc_{l_i} \sin xl_i/m \sin g_i t/T}{l_i \rho(l_i) \cos g_i \pi},$$

where we have made use of the relation

$$\frac{g_i}{T} = \frac{l_i \rho(l_i)}{m}.$$

Evidently

$$(22.3) \qquad y'_t(x, T\pi) = \sum_{1}^{N} c_{l_i} \sin \frac{xl_i}{m}.$$

Suppose there were another solution $y''(x, t)$ consistent with the Dirichlet data and (22). Then

$$(22.4) \qquad y^* = y' - y''$$

vanishes on the boundary and satisfies (22.1). Hence y^* vanishes identically. This implies that $y'(x, t)$ is unique. Thus there is an N-dimensional linear manifold of solutions, determined by

$$(22.5) \qquad \sin \frac{xl_i}{m} \sin \frac{tg_i}{T} \qquad (i = 1, \cdots, N).$$

If we restrict admissible solutions by the requirement

$$(22.6) \qquad \int_0^{T\pi} \int_0^{m\pi} y(x, t) \sin \frac{l_i x}{m} \sin \frac{g_i t}{T} \, dx \, dt = a_i \quad (i = 0, 1, \cdots, N),$$

it is clear that the values of the constants $\{c_{l_i}\}$ are uniquely defined in terms of the constants $\{a_i\}$. That is to say, there is a unique solution given by (22.2) in conformity with the conclusion asserted in Theorem 4′. If some one or all of the constants $\{a_i\}$ are 0, then (22.6) may be replaced by either of the conditions

$$(22.61) \qquad \int_0^{m\pi} y(x, t) \sin \frac{xl_i}{m} \, dx = 0$$

or

$$(22.62) \qquad \int_0^{T\pi} y(x, t) \sin \frac{tg_i}{T} \, dt = 0.$$

THEOREM 6. *If $y(x, t)$ satisfies the hypotheses of Theorem 2 and α and β are in the categories listed in the preceding theorems, then, if a solution does exist, it is either unique or unique up to the linear manifolds occurring in Theorems 3, 4 and 5.*

If we include (22.6) with say $a_i \equiv 0$ in the hypotheses of Theorem 6, then an admissible solution, if it exists, must be unique. The demonstration of Theorem 6 is obvious.

We turn now to questions connected with the existence of solutions. The assigned boundary functions $y(x, 0)$, $y(x, T\pi)$ may be extended to all x values by requiring the extended functions to be odd and periodic of period $2m\pi$. Similarly the extended functions $y(0, t)$, $y(m\pi, t)$ are to be odd and periodic of period $2T\pi$. This is the sense in which the designation *extended boundary functions* is to be understood. Suppose the Dirichlet data are expressible in Fourier sine series with a_l, A_l, b_l, B_l the l-th coefficients in the associated expansions of $y(x, 0)$, $y(x, T\pi)$, $y(0, t)$ and $y(m\pi, t)$ respectively. A formal solution is

$$\begin{aligned} y(x, t) = \sum_{l=1}^{\infty} \Bigg[& \frac{(a_l \sin l\rho(l)(T\pi - t)/m + A_l \sin l\rho(l)t/m) \sin lx/m}{\sin \alpha\pi l\rho(l)} \\ & + \frac{(b_l \sin l\nu(l)(m\pi - x)/T + B_l \sin l\nu(l)x/T) \sin lt/T}{\sin \pi l\nu(l)/\alpha} \Bigg]. \end{aligned} \tag{25}$$

For the existence of a solution of class C^2 expressed in the form (23) it will be sufficient to place such conditions on the extended boundary functions as will yield

$$\left|\frac{a_l}{\sin \alpha\pi l\rho(l)}\right|, \quad \left|\frac{A_l}{\sin \alpha\pi l\rho(l)}\right|, \quad \left|\frac{b_l}{\sin \pi l\nu(l)/\alpha}\right|, \quad \left|\frac{B_l}{\sin \pi l\nu(l)/\alpha}\right| \leqq \frac{N}{l^4} \quad \text{for } l \geqq L. \tag{23.1}$$

We may make use of the classic result that if $f(x) \in C^q$, $q \geqq 0$, and is odd, the coefficients in its Fourier sine expansion, ordered by the l, are $o(l^{-q})$. Lemmas 6, 7, 8 and 10 taken in conjunction with Lemmas 12 and 13 assert that for certain classes of α and β, the terms $|\sin \alpha l\rho(l)|$, $|\sin l\nu(l)/\alpha|$ are dominants for the same negative integral powers of l. Let this negative power of l be $l^{-\mu}$. In order, then, to guarantee the inequality in (23.1), it is merely necessary to take extended data of class $C^{\mu+4}$. In view of our previous theorems it is clear that either the solution (23) is unique or is a member of a linear manifold of solutions determined by the nature of α and β.

With these ideas in mind it is a simple matter to write down a large number of existence theorems, containing uniqueness or quasi-uniqueness assertions, by combining our lemmas and theorems in various ways. It is worth while to state some of the characteristic theorems arising in this way.

THEOREM 7. *If* (a) *α is rational,* (b) *β is irrational,* (c) *the extended boundary functions are of class* C^5, *then there is a unique admissible solution which is actually of class* C^2.

Evidently hypotheses (a) and (b) preclude the existence of solutions of (A).

THEOREM 8. *If* (a) *$\alpha = r/s$ is a rational number,* (b) *$\beta = P/Q$ is a rational number,* (c) *the extended data are of class* C^5, *then an admissible solution exists which is actually of class* C^2. *This solution is either unique or a member of a finite manifold of solutions depending on whether or not a solution of* (A) *exists and the division into various cases in terms of r, s, P and Q is given in Lemma* 3.

Theorem 9. *If* (a) $\alpha = (r/s)^{1/2}$, *where the integers r and s are not both squares,* (b) $\beta = P/Q$, *a rational number,* (c) *the extended data are of class* C^8, *then an admissible solution exists which is actually of class* C^2. *This solution is unique or a member of a manifold of solutions of infinite dimensionality depending on whether or not* (A) *has solutions and the division into cases determined by r, s, P and Q is that of Lemma* 4.

Theorem 10. *If* (a) α *is an algebraic number of degree* $S > 1$, (b) β *is an algebraic number of degree R,* (c) *the extended data is of class* C^{2RS+4}, *then an admissible solution exists which is actually of class* C^2. *This solution is either unique or is unique up to a one-dimensional linear manifold. The distinction between these cases is covered by Lemma* 11.

We proceed to show by an example (cf. (25)) that hypothesis (c) in Theorems 7, 8, 9 and 10 is not necessary. First we establish a lemma of use in the sequel.

Lemma 15. *The series* (a) $\sum_1^\infty (\sin \alpha\pi l\rho(l) \cos lx/m)/l$ *and* (b) $\sum_1^\infty (\sin \alpha l\pi \cdot \cos lx/m)/l$ *have the same convergence properties.*

Evidently the series in question are Fourier cosine series. We denote the series in (a) and (b) by the symbols $S(x)$ and $S'(x)$ respectively. Now

$$|\sin \alpha\pi l\rho(l) - \sin \alpha\pi l| = 2\,|\cos \tfrac{1}{2}\alpha\pi l(1+\rho(l)) \sin \tfrac{1}{2}\alpha\pi l(1-\rho(l))|. \tag{24}$$

Let

$$\delta = \tfrac{1}{2}\alpha\beta\pi.$$

Then, for $l > L$,

$$\frac{1}{2} > \frac{\delta}{l} \geqq \frac{1}{2}\alpha l(1-\rho(l)). \tag{24.1}$$

Thus

$$|\sin \alpha\pi l(1-\rho(l))| < \frac{\delta}{l}, \qquad l > L, \tag{24.2}$$

and it follows from (24) and (24.2) that

$$\sum_p^\infty \left|\frac{\sin \alpha\pi l\rho(l) - \sin \alpha\pi l}{l}\right| \leqq 2\sum_p^\infty \frac{\delta}{l^2} < \frac{2\delta}{p}. \tag{24.21}$$

For $n > p$,

$$\begin{aligned} |S_n(x) - S_p(x)| &\leqq |(S_n(x) - S'_n(x)) - (S_p(x) - S'_p(x))| + |S'_n(x) - S'_p(x)| \\ &\leqq \sum_p^\infty \left|\frac{\sin \alpha\pi l\rho(l) - \sin \alpha\pi l}{l}\right| + \left|\sum_p^n \left(\frac{\sin \alpha\pi l \cos lx/m}{l}\right)\right|. \end{aligned} \tag{24.3}$$

It is well known that $S'(x)$ is a uniformly convergent series in any closed interval in $-m\pi \leqq x \leqq m\pi$ not containing the points

$$x = \pm T\pi. \tag{24.4}$$

Consider such a closed interval J. Then, for arbitrary $\epsilon > 0$, p can be chosen so that for all $n > p$ $(> 4\delta/\epsilon)$ and $x \in J$,

$$\left|\sum_p^n \frac{\sin \alpha\pi l \cos lx/m}{l}\right| < \tfrac{1}{2}\epsilon. \tag{24.5}$$

Hence by (24.21), (24.3) and (24.5)

$$| S_n(x) - S_p(x) | < \epsilon, \qquad n \geqq p > \frac{4\delta}{\epsilon}, \qquad x \in J. \tag{24.6}$$

That is to say, $S_n(x)$ converges uniformly in the interval J.

We show now that the series $S_n(x)$ is not uniformly convergent in any neighborhood U of any of the points singled out in (24.4). We have

$$| S_n(x) - S_p(x) | \geqq | S_n'(x) - S_p'(x) | - \frac{2\delta}{p}. \tag{24.7}$$

Since $S'(x)$ is not uniformly convergent in U, there is a constant C such that for each choice of p there exists an integer n, $n > p$, for which

$$| S_n'(x) - S_p'(x) | \geqq C, \tag{24.8}$$

for some x in U. Hence by (24.7)

$$| S_n(x) - S_p(x) | > \tfrac{1}{2}C, \qquad p > \frac{C}{4\delta}, \tag{24.71}$$

for some integer n, $n > p$, and a point x of U. That is to say, $S(x)$ is not uniformly convergent in U. This completes the proof of the lemma.

Consider the following Dirichlet data:

$$\begin{aligned} y(x, 0) &= \sum \frac{\sin \alpha\pi l \rho(l) \sin lx/m}{l^4}, \\ y(0, t) &= y(m\pi, t) = y(x, T\pi) = 0. \end{aligned} \tag{25}$$

We make use of Lemma 15. Since the coefficients in $S(x)$ are $O(l^{-1})$, $S(x)$ is the Fourier series of a summable function and, as is well known, may therefore be obtained by termwise differentiation of the series for the indefinite integral of this function.[9] Accordingly,

$$\frac{\partial^3 y(x, 0)}{\partial x^3} \sim \frac{1}{m^3} S(x). \tag{25.1}$$

The uniform convergence of $S(x)$ in the complement of neighborhoods of the points in (24.4) guarantees that the left side of (25.1) is continuous except possibly at these points. If $y(x, 0)$ belonged to C^4, then $S(x)$ would perforce be uniformly convergent throughout the period interval, in contradiction to

[9] A. Zygmund, *Trigonometrical Series*, Warsaw, 1935, p. 15.

Lemma 15. Hence $y(x, 0)$ is of class C^3 at most. A solution of the Dirichlet problem, with the data of (25), is

$$y(x, t) = \sum \frac{\sin lx/m \sin l\rho(l)(T\pi - t)/m}{l^4} \tag{25.2}$$

and evidently $y(x, t)$, so defined, is of class C^2. We have, then, exhibited data of class C^3, at most, for which, none the less, a solution of class C^2 exists for *all* values of α and β.

We proceed now to connect the Dirichlet problem with an associated Cauchy problem. For the purposes of this paper the following special lemma is sufficient.

LEMMA 16. *Suppose $Y(0, t) = Y(m\pi, t) = 0$ and such Dirichlet data on $t = 0$, $T\pi$ that a unique admissible solution $y^*(x, t) \in C^2$ exists. Then there is a Cauchy problem with data on $t = 0$, $T\pi \leqq x \leqq (m + T)\pi$ for which the unique solution coincides with $y^*(x, t)$ in the rectangle. The Cauchy data are odd and of period $2m\pi$.*

We take for granted the classical results in the Cauchy problem for normal hyperbolic equations.[10] We extend $y^*(x, t)$ by the convention

$$-y^*(-x, t) = y^*(x, t) = y^*(x \pm 2m\pi, t), \qquad 0 \leqq t \leqq T\pi.$$

There will be no confusion if $y^*(x, t)$ henceforth denotes the extended function. The only possible discontinuities in the second derivative of $y^*(x, t)$ would be jumps at points on the lines $x = \pm N2m\pi$ $(N = 0, 1, \cdots)$. It will appear, however, that such discontinuities do not enter.

Consider the "mixed" data

$$\begin{aligned} &y(0, t) = 0, && 0 \leqq t \leqq T\pi, \\ &y(x, 0) = y^*(x, 0); y_t(x, 0) = y_t^*(x, 0), && 0 \leqq x \leqq T\pi. \end{aligned} \tag{26}$$

A unique solution of (.1) satisfying (26) is determined in the triangle W formed by the lines $x = 0$, $t = 0$, $x + t = T$. This solution must be $y^*(x, t)$.

Consider now the Cauchy data

$$y(x, 0) = y^*(x, 0), \qquad y_t(x, 0) = y_t^*(x, 0). \tag{26.1}$$

Suppose either $y_{xx}^*(x, 0)$ or $y_{tx}^*(x, 0)$ has a jump discontinuity at $x = 0$. Then the admissible solution $\bar{y}(x, t)$ of (.1) subject to (26.1) would not be of class C^2 on the characteristic $x = t$. Now $\bar{y}(x, t)$ automatically vanishes on $x = 0$. Thus $\bar{y}(x, t)$ satisfies (26). Accordingly $\bar{y}(x, t)$ may be identified with $y^*(x, t)$ in W, and also in W_-, the reflection of W in $x = 0$. Since $y^*(x, t)$ is of class C^2 on $x = \pm t$, it follows that $y^*(x, 0)$ and $y_t^*(x, 0)$ are of class C^2 and C^1 respectively for $-T\pi \leqq x \leqq T\pi$. This implies that the solution $\bar{y}(x, t)$ or $y^*(x, t)$ is of class C^2 on the whole of the closed triangle $W + W_-$. In a similar

[10] Courant-Hilbert, *Methoden der mathematischen Physik*, Berlin, 1931, vol. 2, Chapter 3. Hadamard, *Lectures on the Cauchy Problem*, New Haven, 1923.

way it may be shown that the Cauchy data for $m\pi - T\pi \leqq x \leqq m\pi + T\pi$ determines a solution coinciding with $y^*(x, t)$ in the triangle V formed by $t = 0$, $x = m\pi$ and $x - t = (m - T)\pi$.

The trapezoid, resulting from the removal of W and V from the rectangle, supports Cauchy data on the base and Dirichlet data on the two inclined sides (for the inclined sides are sides of W and V). This mixed data problem determines a unique admissible solution of class C^2 in the closed trapezoid.[11] Hence this solution must coincide with $y^*(x, t)$ in the trapezoid. Accordingly, the Cauchy data

$$y(x, 0) = y^*(x, 0), \qquad y_t(x, 0) = y_t^*(x, 0), \qquad (m - T)\pi \leqq x \leqq (m + T)\pi,$$

determine a unique admissible solution which is identical with $y^*(x, t)$. The continuity restrictions of this lemma may be weakened. Moreover, the equivalence asserted in the lemma is valid also for each member of a manifold of solutions and this fact is used in the second proof of Theorem 1 given later.

Suppose now that α and β are such that (A) cannot be satisfied. Let $y(x, t)$ vanish on the two verticals and the base of the rectangle and let the extended function $y(x, T\pi)$ be of class C^2 at least. Assume furthermore that the necessarily unique solution of this Dirichlet problem yields $y_t(x, 0) \in C^1$. Even under less restrictive conditions Lemma 16 implies that the solution may be represented by the Riemann-Picard formula involving the equivalent Cauchy data on $t = 0$,

$$y(x, t) = \frac{1}{2}\int_{x-t}^{x+t} y_t(z, 0)J_0K((z - x)^2 - t^2)^{\frac{1}{2}}\, dz. \tag{27}$$

The Fourier sine coefficients of $y(x, T\pi)$ are denoted by a_l. Therefore

$$\begin{aligned} a_l &= \frac{1}{4m\pi}\int_0^{m\pi}\int_{x-T\pi}^{x+T\pi} y_t(z, 0)J_0K((z - x)^2 - (T\pi)^2)^{\frac{1}{2}} \sin\frac{lx}{m}\, dz\, dx \\ &= \frac{1}{4m\pi}\int_{-T\pi}^{T\pi}\int_{-m\pi}^{m\pi} y_t(x + u, 0) \sin\frac{lx}{m} J_0K(u^2 - (T\pi)^2)^{\frac{1}{2}}\, dx\, dt, \end{aligned} \tag{27.1}$$

where the interchange of integration order, following the substitution $u = z - x$, is justified by the uniform continuity of the integrand in all variables. Let us define $c'_l = c_l \cos lv/m$ by

$$\begin{aligned} 2m\pi c'_l &= \int_{-m\pi}^{m\pi} y_t(x + u, 0) \sin\frac{lx}{m}\, dx \\ &= \int_{-m\pi+u}^{m\pi+u} y_t(v, 0) \sin\frac{l(v - u)}{m}\, dv = \cos\frac{lu}{m}\int_{-m\pi}^{m\pi} y_t(v, 0) \sin\frac{lv}{m}\, dv, \end{aligned}$$

[11] This solution may be built up in the standard way by subdividing the trapezoid into squares or triangles whose boundaries are made up of the characteristics or of segments of the lines $t = 0$ or $t = T\pi$. We continue the solution from subdivision to subdivision. At each stage we have a soluble problem arising from either Cauchy data or Dirichlet data on two intersecting characteristics.

where we have made use of the fact that $y_t(v, 0)$ is odd. Thus c_l is the l-th Fourier sine constant for $y_t(x, 0)$. We have then

$$a_l = \tfrac{1}{2}c_l \int_{-T\pi}^{T\pi} \cos\frac{lu}{m} J_0 K(u^2 - (T\pi)^2)^{\frac{1}{2}}\, du. \tag{27.3}$$

The integral in (27.3) is known, but it is of interest to evaluate it by making use of the considerations of this paper. Indeed, for a fixed value of l, we consider the data

$$Y(x, T\pi) = \sin\frac{lx}{m}, \qquad Y(x, 0) = Y(t, 0) = Y(m\pi, 0) = 0. \tag{27.31}$$

Under our assumption that (A) cannot be satisfied the unique solution, subject to (27.31), is

$$Y(x, t) = \frac{\sin l\rho(l)t/m \sin lx/m}{\sin \alpha\pi l\rho(l)}. \tag{27.32}$$

Therefore

$$Y_t(x, 0) = \frac{l\rho(l)}{m}\,\frac{\sin lx/m}{\sin \alpha\pi l\rho(l)}. \tag{27.33}$$

By (27.31) and (27.32) we have

$$c_l = \frac{l}{m}\,\frac{\rho(l)}{m \sin \alpha\pi l\rho(l)}. \tag{27.34}$$

On substituting these values of a_l and c_l in (27.3) there results[12]

$$\frac{1}{2}\int_{-T\pi}^{T\pi} \cos\frac{ul}{m} J_0 K(u^2 - (T\pi)^2)^{\frac{1}{2}}\, du = \frac{m}{l}\,\frac{\sin \alpha\pi l\rho(l)}{\rho(l)}. \tag{27.4}$$

Hence

$$c_l = a_l \frac{l}{m}\,\frac{\rho(l)}{\sin \alpha\pi l\rho(l)}. \tag{27.5}$$

The result in (27.5) may be obtained by formal operation on an assumed solution of the form (23). Besides the interest of the alternative method based on (27), there is greater generality to the procedure adopted.

We show by an example that some such hypothesis as (c) in Theorems 7, 8, 9, and 10 is required for *sufficiency*. Let us denote by Q' the smallest integer Q for which integers N_l, L and a constant C exist, with the property

$$\tfrac{1}{2} > \left|\alpha\rho(l) - \frac{N_l}{l}\right| > \frac{C}{l^{Q+1}}, \qquad l \geqq L. \tag{28}$$

[12] It is easy to extend (27.4) to all T values.

Suppose further that α and β are so chosen that $Q' \geqq 3$. Then, clearly, there is a constant D such that for an infinite number of pairs N_l, l

$$\left|\alpha\rho(l) - \frac{N_l}{l}\right| < \frac{D}{l^{Q'}}, \tag{28.1}$$

and for these values

$$|\sin(\alpha\rho(l)\pi l - N_l\pi)| = |\sin \alpha l\rho(l)\pi| \leqq \frac{E}{l^{Q'-1}}. \tag{28.11}$$

We consider the Dirichlet problem with $y(x, t)$ vanishing on the base and on the two vertical sides and take $y(x, T\pi)$ as

$$y(x, T\pi) = \sum \frac{\sin lx/m}{l^{Q'}}. \tag{28.2}$$

Manifestly $y(x, T\pi)$ is of class $C^{Q'-2}$. Suppose a solution of class C^1 exists. According to (27.5) and (28.11) we should have

$$c_l \geqq E, \qquad \text{a constant,} \tag{28.3}$$

for the infinite number of l values for which (28.1) holds. The Riemann-Lebesgue lemma precludes the possibility that the sequence $\{c_l\}$ can be Fourier constants. Hence $y_t(x, 0)$ cannot be of class C^1. We have, then, exhibited a case of data of continuity class not inferior to $C^{Q'-2}$ for which no solution of class C^2 exists.

We now give an alternative demonstration of the sufficiency part of Theorem 1 when the solution is required to be of class C^2. For vanishing Dirichlet data, any admissible solution of class C^2 is given by (27). Moreover, it follows from (27.5) that c_l vanishes unless

$$\sin \alpha\pi l\rho(l) = 0. \tag{29}$$

The satisfaction of the relation in (29) would mean the existence of a solution of (A), but this has been ruled out by the conditions of Theorem 1. Hence

$$c_l = 0 \qquad (l = 1, 2, \cdots). \tag{29.1}$$

Since the functions $\{\sin lx/m\}$ are complete in the space of continuous functions defined on $0 \leqq x \leqq m\pi$, it follows that

$$y_t(x, 0) = 0. \tag{29.11}$$

Thus the right side of (27.5) is identically 0, whence the admissible solution must be

$$y(x, t) = 0. \tag{29.2}$$

This completes the proof.

As a by-product of our analysis we derive some results on closure property of sets of functions associated naturally with the Dirichlet problem. Denote

by $D(-M\pi, M\pi)$ the space of odd continuous functions on $-M\pi \leqq z \leqq M\pi$ with the norm $\|f(z)\| = \max_{0\leqq|z|\leqq M\pi} |f(z)|$.

THEOREM 11. *The sequences* $\{\sin lt\rho(l)/m\}$ *and* $\{\sin lt\nu(l)/T\}$ *are closed in* $D(-m\pi, m\pi)$ *and* $D(-T\pi, T\pi)$ *respectively and minimal. Any function in the domains cited which has, besides, a continuous derivative admits a uniformly convergent expansion (in terms of the functions above) with coefficients identical except for a factor* $\rho(l)$ $(\nu(l))$ *with those in the Fourier sine series expansion of an associated function.*[13]

Consider the following Cauchy data

$$y_t(x, 0) = -y_t(-x, 0) \in C^p,\ p \geqq 1; \qquad y(x, 0) = 0. \tag{30}$$

A unique solution $\bar{y}(x, t) \in C^{p+1}$, odd in x and t, is determined in the square bounded by the four characteristics

$$(\pm)x(\pm)t = M\pi. \tag{30.1}$$

In particular $\bar{y}(0, t)$ is zero. Thus the data of (30) imply

$$\begin{aligned} y(0, t) &= y(x, 0) = 0, \\ y_x(0, t) &= -y_x(0, -t) = \bar{y}_x(0, t). \end{aligned} \tag{30.2}$$

If the functions in (30.2) are considered as the assigned data, it is well known that the same solution $(\bar{y}(x, t))$ is determined in the square defined by (30.1). In brief, the functions in (30) and (30.1) are related by a transformation determined by (.1). It will appear that the closure property of sequences of functions is preserved. Perhaps the sole advantage of a transformation viewpoint here and in kindred theorems is indicated by the last sentence of the theorem. The usual (and more powerful) methods of establishing closure are not concerned with the *practical* determination of the coefficients cited.

We proceed to some explicit formulas. It is convenient to write

$$G(z, t) = \frac{1}{2}\frac{\partial}{\partial z} J_0K(z^2 - t^2)^{\frac{1}{2}}. \tag{30.3}$$

Since $J_0K(z^2 - t^2)^{\frac{1}{2}}$ is an entire function in z and t, it is easily established that $G(z, t)$ and the resolvent kernel $H(z, t)$, also, are entire in z and t. The solution of the Cauchy problem with the data of (30) is given by (27). For $y_t(z, 0)$ the usual algorithm for differentiating an integral containing a parameter is valid. Hence, since $J_0(0) = 1$,

$$\begin{aligned} y_x(0, t) &= \frac{1}{2}[y_t(t, 0) - y_t(-t, 0)] \\ &\quad + \frac{1}{2}\left[\int_{x-t}^{x+t} y_t(z, 0)\frac{\partial}{\partial x} J_0K((z - x)^2 - t^2)^{\frac{1}{2}}\,dz\right]_{x=0} \\ &= y_t(t, 0) - \int_{-t}^{t} y_t(z, 0)G(z, t)\,dz. \end{aligned} \tag{30.4}$$

[13] Theorem 11 may be used in our original proof of Theorem 1.

If $y_t(z, 0) \in C^p$, $p \geqq 1$, we may integrate by parts. Thus

$$y_x(0, t) = y_t(0, t) - \frac{1}{2}[y_t(z, 0)J_0K(z^2 - t^2)^{\frac{1}{2}}]\Big|_{-t}^{t} + \frac{1}{2}\int_{-t}^{t} y_{tz}(z, 0)J_0K(z^2 - t^2)^{\frac{1}{2}}\,dz$$

$$(30.41) \qquad = \frac{1}{2}\int_{-t}^{t} y_{tz}(z, 0)J_0K(z^2 - t^2)^{\frac{1}{2}}\,dz.$$

(It is, perhaps, well to comment that the first argument in $y_t(\ )$ and its derivatives refers to values on the *x-axis*. Hence, for example, $y_t(t, 0)$ is $y_t(x, 0)$ for the special value $x = t$.)

The classical theory of the Volterra integral equation (30.4) with a kernel analytic in both arguments asserts the existence of a unique odd solution $y_t(t, 0)$ of the same continuity class as $y_x(0, t)$, namely,

$$(30.5) \qquad y_t(t, 0) = y_x(0, t) + \int_{-t}^{t} y_x(0, z)H(z, t)\,dz.$$

Moreover, (30.4) and (30.5) are valid if the functions are merely odd and continuous, i.e., $D(-M\pi, M\pi)$ is taken into itself by (30.4) or (30.5).

A closed form for the resolvent kernel $H(z, t)$ follows at once from the observation that if x and t are interchanged in (.1), the effect is tantamount to replacing K by $-K$. Hence

$$H(z, t) = -\frac{1}{2}\frac{\partial}{\partial t} J_0K(z^2 - t^2)^{\frac{1}{2}}.$$

Thus the solution of the Cauchy problem with the data of (30.2) provides the inversion of (30.4) or, if $y_x(0, t) \in C^p$, $p \geqq 1$, of (30.41), i.e.,

$$(30.42) \qquad y_t(u, 0) = y_x(0, u) + \frac{1}{2}\int_{-u}^{u} y_x(0, z)\frac{\partial}{\partial z} J_0K(u^2 - z^2)^{\frac{1}{2}}\,dz,$$

$$(30.43) \qquad y_t(u, 0) = \frac{1}{2}\int_{-u}^{u} y_{tz}(0, z)J_0K(u^2 - z^2)^{\frac{1}{2}}\,dz.$$

We consider only $\{\sin lt\rho(l)/m\}$ since the argument for $\{\sin lt\nu(l)/T\}$ is similar.[14] Suppose $f(t) \in D(-m\pi, m\pi)$. We identify it with $y_x(0, t)$ and use (30.5) to define the function $y_t(t, 0) \in D(-m\pi, m\pi)$. Let $\{\phi_n(t)\}$ be closed in $D(-m\pi, m\pi)$ and let $\psi_n(t)$ correspond to $\phi_n(t)$ by (30.5). Suppose

$$(30.6) \qquad \left\| y_t(t, 0) - \sum_{1}^{N} c_n\phi_n(t) \right\| < \epsilon.$$

[14] In view of (30.42) and (30.43) we identify $f(t)$ with $y_t(t, 0)$ and use now, in place of (27.4), the result, easily subsumed under the Gegenbauer finite integral,

$$\frac{1}{2}\int_{-u}^{u} \cos\frac{lz}{T} J_0K(u^2 - z^2)^{\frac{1}{2}}\,dz = \frac{T}{l}\,\frac{\sin \pi l\nu(l)/\alpha}{\nu(l)}.$$

G. N. Watson, *Theory of Bessel Functions*, 1922, p. 378.

If B bounds $H(x, t)$ in the square defined in (30.1), it follows at once that

$$\left\| f(t) - \sum_1^N c_n \psi_n(t) \right\| \leqq \epsilon(1 + B). \tag{30.7}$$

Thus $\{\psi_n(t)\}$ is closed in $D(-m\pi, m\pi)$ and the coefficients c_n for the approximation of $f(t)$ may be taken as those occurring in (30.6). The result stated in the theorem is the special case of the foregoing in which $\phi_n(t)$ is taken as $\sin nt/m$. On making use of (30.42) it is evident that if $\{\phi_n(t)\}$ is minimal, then $\{\psi_n(t)\}$ is minimal.

If $f(t) \in C^2$ then $y_t(t, 0) \in C^2$ and the coefficients $c_l = O(l^{-3})$ may be taken as the Fourier sine coefficients in the expansion of $y_t(t, 0)$. Indeed the Fourier sine series for $y_t(t, 0)$ may be differentiated term by term to give a uniformly convergent series, viz.,

$$\begin{aligned} y_t(z, 0) &= \sum c_l \sin \frac{l}{m} z, \\ y_{tz}(z, 0) &= \sum c_l \frac{l}{m} \cos \frac{l}{m} z. \end{aligned} \tag{30.8}$$

By (30.41)

$$y_z(0, t) = \frac{1}{2} \int_{-t}^{t} \sum c_l \frac{l}{m} \cos \frac{l}{m} z J_0 K(z^2 - t^2)^{\frac{1}{2}} \, dz.$$

The order of integration and summation may be interchanged by the uniform convergence of the series in the integrand and we get

$$f(t) = \sum c_l \frac{\sin l\rho(l)t/m}{\rho(l)}. \tag{30.9}$$

Since $\rho(l) \uparrow 1$, the series in (30.9) is uniformly convergent, whence it follows that the closed set $\{(\sin l\rho(l)t/m)/\rho(l)\}$ may be used with the Fourier coefficients c_l formed for the associated function $y_t(t, 0)$. On using (30.4) instead of (30.41) this remark is established for $f(t) \in C^1$. The closure of $\{\sin l\rho(l)t/m\}$ in $D(-m\pi, m\pi)$ follows from the fact that the functions of class C^2 are everywhere dense in $D(-m\pi, m\pi)$.

The results obtained so far admit of considerable generalization. Consider the contravariant tensors $g^{ij}(x^1, x^2)$ and $h^i(x^1, x^2)$ $(i, j = 1, 2)$ and the scalar $c(x^1, x^2)$. All these quantities may, for simplicity, be considered analytic in x^1 and x^2 in a region E. We assume that the associated covariant tensor $g_{ij}(x^1, x^2)$ has analytic components and that[15]

$$g_{ij} \, dx^i \, dx^j \tag{31}$$

is a non-singular indefinite form in E. Furthermore suppose

$$\frac{\partial h^1}{\partial x^2} = \frac{\partial h^2}{\partial x^1}. \tag{31.1}$$

[15] The usual dummy script summation convention of tensor analysis is used here.

We define $\bar{L}(\bar{y})$ as

$$\bar{L}(\bar{y}) = g^{ij}\,\frac{\partial^2 \bar{y}}{\partial x^i \partial x^j} + h^i\,\frac{\partial y}{\partial x^i} + cy. \tag{31.2}$$

It is well known[16] that non-identically vanishing functions $f^1(x^1, x^2)$ and $f^2(x^1, x^2)$ can be found such that a suitable transformation

$$x^i = x^i(x, t) \qquad (i = 1, 2) \tag{31.3}$$

yields

$$\frac{f^2}{f^1}\,\bar{L}(\bar{y}f^1) = L(y), \tag{31.4}$$

where $L(y)$ is the expression already defined in (.1) and $\bar{y} = y(x(x^1, x^2), t(x^1, x^2))$. Hence solutions of (.1) are correlated with solutions of

$$\bar{L}(f^1\bar{y}) = 0. \tag{31.5}$$

Suppose (31.3) defines a homeomorphism between a set in the x, t-plane containing the rectangle and the set E. The boundary of the image of the rectangle is denoted by $\bar{\Gamma}$. If the functions $x^1(x, t)$ and $x^2(x, t)$ are of sufficiently high continuity class (for instance, superior to that required of the data in our various preceding theorems), it is plain that questions of uniqueness and existence of solutions for the Dirichlet problem connected with (31.5) and data on $\bar{\Gamma}$ may be treated by transformation to the canonical situation (.1) and (.2).

An interesting specialization of these general remarks is concerned with the determination of classes of contours for which the results of our paper remain valid for (.1). Here we are concerned with the transformations leaving (.1) invariant. If $K \neq 0$ it is apparent that the transformations must leave invariant the quadratic form

$$dx^2 - dt^2.$$

Thus we are led to the Lorentz transformation

$$\bar{x} = \alpha_1 x + \alpha_2 t, \qquad \bar{t} = \alpha_2 x + \alpha_1 t, \qquad 1 = \alpha_1^2 - \alpha_2^2\,, \tag{32}$$

$$d\bar{x}^2 - d\bar{t}^2 = dx^2 - dt^2. \tag{32.1}$$

Since the characteristics, namely, the lines of slopes ± 1, are left invariant, it is clear that the rectangle goes into a parallelogram, with the sides $x = 0$, $t = 0$ transforming into lines *equally inclined to the characteristic directions.* Moreover, since

$$\bar{x}^2 + \bar{t}^2 = (\alpha_1^2 + \alpha_2^2)(x^2 + t^2) + 4\alpha_1\alpha_2 tx, \tag{32.2}$$

[16] E. Cotton, *Sur les invariants différentiels de quelques équations linéaires aux dérivées partielles du second ordre*, Annales Scientifiques de l'Ecole Normale Supérieure, vol. 38(1900), p. 211.

we can easily show, by taking $x = 0$, or $t = 0$, that the ratio of the lengths of the corresponding sides of the parallelogram remains α. We replace normal derivatives by co-normal derivatives and the new and old solutions are related by

$$\bar{y}(\bar{x}, \bar{t}) = y(x(\bar{x}, \bar{t}), t(\bar{x}, \bar{t})). \tag{32.3}$$

For completeness we consider the case $K = 0$ also. The following remarks indicate the nature of the extensions of previous work on the vibrating string equation. The class of transformations leaving (.1) invariant is now the class of conformal transformations defined by

$$d\bar{x}^2 - d\bar{t}^2 = \rho(x, t)(dx^2 - dt^2). \tag{33}$$

It is well known that the general conformal transformation in the sense of (33) is given by

$$\bar{x} = F_1(x + t) + F_2(x - t), \qquad \bar{t} = F_1(x + t) - F_2(x - t). \tag{33.1}$$

Observe that the characteristics go into characteristics and that the Jacobian of the transformation is $\rho(x, t)$. If we can ensure that admissible solutions correspond in the two sets of variables, the theory developed for the rectangle *can be taken over bodily*. It is sufficient for this to require as before that the functions on the right side of (33.1) are of continuity class superior to that required for the data in existence theorems, and that (33.1) defines a homeomorphism on a region containing the rectangle.

UNIVERSITY OF ILLINOIS.

LA LOI DE JORDAN-HÖLDER DANS LES HYPERGROUPES ET LES SUITES GENERATRICES DES CORPS DE NOMBRES $\mathfrak{P}$-ADIQUES

PAR MARC KRASNER

Chapitre II

Suites génératrices des corps de nombres $\mathfrak{P}$-adiques

8. A partir d'ici il ne s'agira que d'hypergroupes$_D$. Dans ces hypergroupes l'infra-invariance se confond avec la semi-invariance à droite, qui sera dite dans la suite du travail semi-invariance tout court. Et la semi-invariance à gauche coïncide avec l'invariance. On dira *quotient* tout court au lieu de dire quotient droit et on le notera H/h au lieu de $H/h^{(\alpha)}$.

L'application qui suit de la théorie précédente à la théorie des corps de nombres $\mathfrak{P}$-adiques montre, en même temps, qu'il existe dans les hypergroupes de très nombreux sous-hypergroupes semi-invariants (à droite) qui ne sont pas invariants et qui y joignent un rôle important. Corrélativement, on peut conclure qu'il existe dans les hypergroupes des sous-hypergroupes semi-invariants à gauche, mais non invariants, importants.

Ceci montre, d'ailleurs, que l'étude précédente a un intérêt qui n'est pas seulement formel.

D'autre part, comme je l'avais déjà mentionné, j'ai trouvé un exemple de sous-groupe infra-invariant, mais non invariant, d'un groupe. Et, un tel sous-groupe ne peut pas être semi-invariant.

9. **Hypergroupe de Galois. Lois de composition induites.** Soit k un corps et soient K' et $K \supseteq K'$ deux surcorps de k de degré fini. Soit σ un isomorphisme de K/k (avec un de ses conjugués). Il existe un et un seul isomorphisme σ' de K'/k tel que pour tout $\alpha' \in K'$ on ait $\sigma'\alpha' = \sigma\alpha'$. Il sera dit *correspondant de σ* dans K' (corr.$_{K'}$ σ) et le signe corr.$_{K'}$ sera regardé comme une fonction définie dans l'ensemble $G_{K/k}$ de tous les isomorphismes de K/k avec ses conjugués. On sait que corr.$_{K'}$ $(G_{K/k}) = G_{K'/k}$. Ceci permet de définir dans $G_{K'/k}$ une fonction gen.$_K$ σ' ($\sigma' \in G_{K'/k}$) par la condition que gen.$_K$ σ' soit l'ensemble de tous les $\sigma \in G_{K/k}$ tels que corr.$_{K'}$ $\sigma = \sigma'$.

Soient σ_1, σ_2 deux éléments de $G_{K/k}$. Soit α un élément primitif de K/k. Désignons par $\sigma_1\sigma_2$ l'ensemble de tous les $\sigma \in G_{K/k}$ tels que toute relation rationnelle $\pi(\alpha, \sigma_2\alpha) = 0$ dans k entre α et $\sigma_2\alpha$ reste encore vraie pour $\sigma_1\alpha$, $\sigma\alpha$ (c'est-à-dire $\pi(\sigma_1\alpha, \sigma\alpha) = 0$). Visiblement, si K^*/k est un surcorps galoisien de

Received January 10, 1940. The first chapter of this paper appeared in this Journal, vol. 6(1940), pp. 120–140. The notation [K1, 26] refers to page 26 of the article designated as K1 in the bibliography at the end of the present chapter.

K/k, $\sigma_1\sigma_2 = \text{corr.}_K\,[\text{gen.}_{K^*}\,\sigma_1\cdot\text{gen.}_{K^*}\,\sigma_2]$ donc ne dépend pas du choix de α. $G_{K/k}$ organisé par cette loi de composition est un hypergroupe$_D$ fini, isomorphe à $G_{K^*/k}/G_{K^*/K}$, qui s'appellera *hypergroupe de Galois* (ou simplement hypergroupe) *de* K/k. Les propriétés des hypergroupes$_D$ montrent que le théorème fondamental de la théorie de Galois peut se formuler ainsi:

Si $K \supset K' \supset k$, $G_{K/K'}$ *est un sous-hypergroupe de* $G_{K/k}$; *si* h *est un sous-hypergroupe de* $G_{K/k}$, *il existe un et un seul corps* K', $K \supset K' \supset k$, *tel que* $G_{K/K'} = h$; $G_{K'/k} \simeq G_{K/k}/G_{K/K'}$, *et cet isomorphisme se réalise par la correspondance* $\sigma \to \text{corr.}_{K'}\,\sigma$ $(\sigma \in G_{K/k})$ [K1, 31–34; K2, 80].

Une chaîne de corps

$$K = Q^{(0)} \supset Q^{(1)} \supset \cdots \supset Q^{(s)} = k \tag{1}$$

s'appelle *suite génératrice de* K/k si, pour tout $i = 1, 2, \cdots, s$, $Q^{(i-1)}/Q^{(i)}$ est une extension primitive. s s'appellera *longueur* de la suite (1), la suite $(Q^{(0)}:Q^{(1)})$, $(Q^{(1)}:Q^{(2)}), \cdots, (Q^{(s-1)}:Q^{(s)})$ s'appellera sa *suite de degrés*, et la suite G_{τ_i}, où $\tau_i = Q^{(i-1)}/Q^{(i)}$ $(i = 1, 2, \cdots, s)$ s'appellera sa *suite d'hypergroupes*. Visiblement, la suite (1) est une suite génératrice de K/k si, et seulement si, $G_{K/Q^{(s)}}$, $G_{K/Q^{(s-1)}}, \cdots, G_{K/Q^{(1)}}$, $G_{K/Q^{(0)}}$ est une suite génératrice de $G_{K/k}$, et dans ce cas les longueurs de deux suites sont égales, la suite de degrés de la suite (1) coïncide (à l'ordre près) avec la suite d'indices de la seconde suite, et la suite d'hypergroupes de la suite (1) avec l'ensemble des quotients de la seconde suite (à l'ordre et isomorphisme près).

Soit $\mathfrak{L}$ un surcorps de k. Si α est un élément primitif de K/k, il est aussi élément primitif de $K\mathfrak{L}/\mathfrak{L}$. Si $\sigma_1, \sigma_2 \in \text{corr.}_K\, G_{K\mathfrak{L}/\mathfrak{L}}$, posons $(\sigma_1\sigma_2)$ égal à l'ensemble de tous les $\sigma \in G_{K/k}$ tels que toute relation rationnelle $\pi(x, y) = 0$ dans $\mathfrak{L}$ qui est juste pour $x = \alpha$, $y = \sigma_2\alpha$ le soit pour $x = \sigma_1\alpha$, $y = \sigma\alpha$. Alors $\text{corr.}_K\, G_{K\mathfrak{L}/\mathfrak{L}}$ sera organisé en un hypergroupe$_D$ isomorphe à $G_{K\mathfrak{L}/\mathfrak{L}}$. Si K^*/k est un surcorps galoisien de $K\mathfrak{L}/k$, et si $U^* = G_{K^*/\mathfrak{L}}$, manifestement, $\text{corr.}_K\, G_{K\mathfrak{L}/\mathfrak{L}} = \text{corr.}_K\, U^*$ et

$$(\sigma_1\sigma_2)_{\mathfrak{L}} = \text{corr.}_K\,[(\text{gen.}_{K^*}\,\sigma_1 \cap U^*)(\text{gen.}_{K^*}\,\sigma_2 \cap U^*)].$$

On appellera la loi de composition $(\sigma_1\sigma_2)_{\mathfrak{L}}$ de l'hypergroupe $U = \text{corr.}_K\, U^*$ *loi de composition induite par* U^*. Cet hypergroupe sera noté $U^{(U^*)}$. On écrira aussi, au lieu de $(\sigma_1\sigma_2)_{\mathfrak{L}}$, $(\sigma_1\sigma_2)^{(U^*)}$ [K1, 34–35].

Il est visible que si $U^{*\prime}$ est un sous-groupe de U^*, et si $\sigma_1, \sigma_2 \in \text{corr.}_K\, U^{*\prime}$, $(\sigma_1\sigma_2)^{(U^{*\prime})} \subseteq (\sigma_1\sigma_2)^{(U^*)}$. Si $W \subseteq \text{corr.}_K\, U^{*\prime}$ est un sous-hypergroupe de $(\text{corr.}_K\, U^*)^{(U^*)}$ donc $(WW)^{(U^*)} \subseteq W$, on a $(WW)^{(U^{*\prime})} \subseteq (WW)^{(U^*)} \subseteq W$, et W est aussi un sous-hypergroupe de $(\text{corr.}_K\, U^{*\prime})^{(U^{*\prime})}$. *Et si*, de plus, W *est semi-invariant dans* $(\text{corr.}_K\, U^{*\prime})^{(U^{*\prime})}$, *il l'est dans* $(\text{corr.}_K\, U^{*\prime})^{(U^*)}$. En effet, si $\sigma \in \text{corr.}_K\, U^{*\prime}$, on a $(W\sigma)^{(U^*)} \supseteq (W\sigma)^{(U^{*\prime})}$ et $(\sigma W)^{(U^*)} \supseteq (\sigma W)^{(U^{*\prime})}$. Or, $(\sigma W)^{(U^*)}$ et $(\sigma W)^{(U^{*\prime})}$ étant des classes à droite dans deux hypergroupes$_D$ finis suivant un même sous-hypergroupe W, ils ont le même nombre d'éléments égal à celui de W; donc $(\sigma W)^{(U^{*\prime})} = (\sigma W)^{(U^*)}$. Donc, si $(W\sigma)^{(U^{*\prime})} \supseteq (\sigma W)^{(U^{*\prime})}$, on a aussi $(W\sigma)^{(U^*)} \supseteq (\sigma W)^{(U^*)}$, et tout est prouvé.

10. Suites génératrices des corps $\mathfrak{P}$-adiques. Soient k un corps de nombres $\mathfrak{p}$-adiques et K un surcorps algébrique de degré fini de k. Soient $\mathfrak{P}$, $\mathfrak{p} = \mathfrak{P}^e$ et $p = \mathfrak{P}^E = \mathfrak{p}^{e_0}$ respectivement l'idéal premier de K, celui de k et le premier rationnel qu'ils divisent. On désignera par $(K:k)$ le degré de K/k. f désignera le degré de $\mathfrak{P}$ dans K/k. α étant un élément algébrique par rapport à K, $\omega(\alpha)$ notera l'ordre de α pour $\mathfrak{P}$.

J'avais construit dans mes travaux cités K1 et K2 une théorie de la ramification dans les corps non-galoisiens de nombres algébriques et de nombres $\mathfrak{P}$-adiques qui généralise celle des corps galoisiens de M. Hilbert.

J'en dois rappeler ici quelques définitions et résultats:

Soit $\sigma \in G_{K/k}$. Le minimum de $\omega(\sigma\alpha - \alpha) - 1$, où α parcourt tous les entiers du corps K, s'appelle *nombre caractéristique de* σ et se note $v(\sigma)$ [K1, 39; K2, 81].

Soient

$$v_0, v_1, v_2, \cdots, v_{m-1}, v_m = +\infty$$

toutes les valeurs positives que peut prendre $v(\sigma)$. Posons, de plus, $v_{-2} = -1$, $v_{-1} = 0$. v_q $(q = -2, -1, 0, 1, \cdots, m)$ s'appelle *q-ième nombre de ramification de K/k.*[44] *v_q est une fraction rationnelle dont le dénominateur est*, comme je l'avais montré [K1, 73–76; K2, 93], *premier à p*. L'ensemble V_q de tous les $\sigma \in G_{K/k}$ tels que $v(\sigma) \geqq v_q$ s'appelle *ensemble de ramification d'ordre q de K/k* (en particulier, V_{-1} se désigne aussi par T et s'appelle aussi *ensemble d'inertie de* K/k, et V_0 se note V et s'appelle aussi *ensemble de ramification de* K/k) [K1, 39; K2, 83]. J'avais prouvé [K1, 43; K2, 82] que *tous les V_q $(q = -2, -1, 0, \cdots, m)$ sont des sous-hypergroupes de l'hypergroupe de Galois $G_{K/k}$ de K/k.* Donc, pour tout $q = -2, -1, 0, \cdots, m$, *il existe un corps K_q, $k \subseteq K_q \subseteq K$, tel que $G_{K/K_q} = V_q$.* K_q/k s'appelle *q-ième corps de ramification* de K/k (en particulier K_{-1}/k s'appelle *corps d'inertie* et K_0/k *corps de ramification* de K/k) [K1, 54; K2, 83]. K_{-1}/k est le plus petit sous-corps Q/k de K/k tel que K/Q soit complètement ramifié, et est le plus grand sous-corps Q/k de K/k non-ramifié [K1, 55; K2, 95].

Soit n_q le nombre d'éléments de V_q. $n_{-2} = fe$, $n_{-1} = e$, *n_0 est la contribution de p dans e, les n_q $(0 \leqq q \leqq m)$ sont des puissances de p, décroissantes avec q* [K1, 43, 46, 52; K2, 85, 90–91]. K^*/k étant un surcorps galoisien arbitraire de K/k, T^* et V^* étant resp. groupe d'inertie et groupe de ramification de K^*/k, on a $\text{corr.}_K\, T^* = T$ et $\text{corr.}_K\, V^* = V$ [K1, 46–47; K2, 86, 89]. Donc les hypergroupes $T^{(T^*)}$ et $V^{(V^*)}$ peuvent être définis. *$G_{K/k}/T$ est un groupe cyclique d'ordre f, et T est invariant dans $G_{K/k}$* [K1, 44–45; K2, 84–85]. *$V^{(T^*)}$ est invariant dans $T^{(T^*)}$ et $T^{(T^*)}/V^{(T^*)}$ est un groupe cyclique d'ordre premier à p* [K1, 48]. *La suite d'indices d'une suite génératrice de l'hypergroupe T/V ne contient que des nombres premiers* [K1, 50, 68; K2, 87–88]. *$V_{q+1}^{(V^*)}$ $(q = 0, 1, \cdots, m-1)$ est invariant dans $V_q^{(V^*)}$ et $V_q^{(V^*)}/V_{q+1}^{(V^*)}$ est un groupe abélien du type $(p, p, \cdots, p)$* [K1, 51–52].

[44] Il est dit nombre de ramification *impropre* si $q = -2, -1$ ou m, et *propre* dans tous les autres cas [K1, 39; K2, 83].

Si K/k est galoisien, V_q ($q = 0, 1, \cdots, m$) est un sousgroupe invariant de $V_{-2} = G_{K/k}$ [O4, 653; K1, 76].[45]

Pour que K/k soit primitif il faut que ou bien $(K:k)$ soit premier, ou bien que K/k soit complètement ramifié de degré puissance de p et n'ait qu'un seul nombre de ramification propre $v = v_0$ [K1, 58–59; K2, 94].

Si $\pi \in K$ est d'ordre 1 pour $\mathfrak{P}$, et si $\sigma \in T$, $\omega(\sigma\pi - \pi) = 1 + v(\sigma)$ [K1, 40–41; K2, 82]. *Si $K_0 = k$, si $K_{-1} = K$ et si δ désigne le dénominateur du nombre de ramification propre $v = v_0$ de K/k il existe un corps non-ramifié $\mathfrak{L}$ tel que tous les conjugués $\sigma\pi$ ($\sigma \in G_{K/k}$) de π par rapport à k soient contenus dans le corps $\mathfrak{L}' = \mathfrak{L}(\pi^{1/\delta})$, obtenu par l'adjonction de $\pi^{1/\delta}$ à $\mathfrak{L}$* [K1, 61; K2, 97, 189].

J'avais prouvé [K1, 63, 70] que si K^*/k est un surcorps (qu'on supposera ici galoisien) de K/k, et si V_q^* désigne son groupe de ramification d'ordre q, il existe une suite (finie) d'entiers croissants et non-négatifs $i_0, i_1, \cdots, i_{m-1}, i_m$ telle que, pour tout q ($0 \leqq q \leqq m$), on ait $\text{gen.}_{K^*} V_q = V_{i_q}^* \cdot G_{K^*/K}$.

LEMME 11. *V_q est semi-invariant dans V ($q = 0, 1, \cdots, m$).*

Démonstration. Soit K^*/k un surcorps galoisien de K/k. Soit $\sigma \in G_{K/k}$, et soit $\sigma^* \in \text{gen.}_{K^*} \sigma$. Alors $\text{gen.}_{K^*} \sigma = \sigma^* G_{K^*/K}$. On a

$$\begin{aligned}\text{gen.}_{K^*} V_q\sigma = \text{gen.}_{K^*} V_q \cdot \text{gen.}_{K^*} \sigma &= V_{i_q}^* G_{K^*/K} \sigma^* G_{K^*/K} \\ &\supseteq V_{i_q}^* \sigma^* G_{K^*/K} = V_{i_q}^* \sigma^* G_{K^*/K} \cdot G_{K^*/K}.\end{aligned}$$

Or, $V_{i_q}^*$ étant invariant dans $G_{K^*/k}$, on a

$$V_{i_q}^* \sigma^* G_{K^*/K} = \sigma^* G_{K^*/K} V_{i_q}^*,$$

donc

$$\text{gen.}_{K^*} V_q\sigma \supseteq \sigma^* G_{K^*/K} \cdot V_{i_q}^* G_{K^*/K} = \text{gen.}_{K^*} \sigma \cdot \text{gen.}_{K^*} V_q = \text{gen.}_{K^*} \sigma V_q,$$

et

$$V_q\sigma \supseteq \sigma V_q.$$

LEMME 12. *K_0 étant égal à k, si $K \supseteq K' \supseteq K'' \supseteq k$, et si K/K' et K''/k sont primitifs, on a pour tout $\sigma \in G_{K/k}$,*

$$G_{K/K''}\sigma \supseteq \sigma G_{K/K'}.$$

Démonstration. Soit i le plus petit nombre tel que $K_i \supseteq K''$. Supposons que la proposition soit juste si l'on remplace K par K_i et K' par $L_i = K' \cap K_i$. ($G_{K/K_i} = V_i$ étant semi-invariant dans $G_{K/k} = V$ et tout sous-hypergroupe de $G_{K/k}$ étant clos, on voit que $G_{K_i/L_i} \simeq G_{K/L_i}/G_{K/K_i} = G_{K/K'}V_i/V_i \simeq G_{K/K'}/(G_{K/K'} \cap V_i)$ et, s'il n'existe aucun hypergroupe entre $G_{K/K'}$ et $\{1_K\}$, où 1_K désigne l'isomorphisme identique de K, il n'en existe aucun, à fortiori, entre G_{K_i/L_i} et $\{1_{K_i}\}$. Donc K_i/L_i est primitif.) Alors, pour tout $\sigma \in G_{K/k}$ on a,

[45] Ce dernier résultat est dû à M. A. Speiser, *Zerlegungsgruppe*, Journal für die reine und angewandte Mathematik, vol. 149(1919), pp. 174–188; voir pp. 175–177.

puisque gen.$_K$ $G_{K_i/L_i} = G_{K/L_i} = G_{K/K'}V_i$, et puisque tout sous-hypergroupe de $G_{K/k}$ est clos,

$$\sigma V_i G_{K/K'} V_i \subseteq G_{K/K''}\sigma V_i \subseteq G_{K/K''}V_i\sigma = G_{K/K''}\sigma.$$

Donc, puisque $\sigma G_{K/K'} = \sigma \cdot 1_K G_{K/K'} 1_K \subseteq \sigma V_i G_{K/K'} V_i$, on voit que la proposition serait juste pour les corps K, K', K'', k eux-mêmes. Donc on peut supposer dans la démonstration $i = m$. Mais alors $K_{m-1} \cap K'' \neq K''$, donc $K_{m-1} \cap K'' = k$. Donc, $G_{K/k} = G_{K/K''}V_{m-1}$.

Supposons qu'on ait prouvé, pour tout $\sigma \in V_{m-1}$, l'inégalité $\sigma G_{K/K'} \subseteq G_{K/K''}\sigma$. Soit σ un élément quelconque de $G_{K/k}$. En vertu de ce qui précède, il existe un $\sigma' \in V_{m-1}$ tel que $\sigma \in G_{K/K''}\sigma'$ et, en vertu de la réversibilité de $G_{K/K''}$, $G_{K/K''}\sigma' = G_{K/K''}\sigma$. On a $\sigma G_{K/K'} \subseteq G_{K/K''}\sigma' G_{K/K'} \subseteq G_{K/K''}G_{K/K''}\sigma' = G_{K/K''}\sigma' = G_{K/K''}\sigma$, et la proposition serait prouvée.

En particulier, puisque $G_{K/K''} \supseteq G_{K/K'}$, il suffirait, pour prouver la proposition, de montrer que, pour tout $\sigma \in V_{m-1}$, on a $\sigma G_{K/K'} \subseteq G_{K/K'}\sigma$; c'est-à-dire, puisque, V_{m-1} étant semi-invariant dans $V = G_{K/k}$, $G_{K/K'}V_{m-1}$ est un hypergroupe, il suffirait de prouver que $G_{K/K'}$ est semi-invariant dans cet hypergroupe. En vertu de §9, il suffirait de prouver ceci pour la loi de composition induite par V^*, où V^* est le groupe de ramification d'un surcorps galoisien quelconque K^*/k de K/k. Nous le ferons en montrant que $(G_{K/K'}V_{m-1})^{(V^*)}$ est un groupe abélien.

En effet, K/K_{m-1} est complètement ramifié de degré puissance de p et n'a qu'un seul nombre de ramification propre v_{m-1}, le plus grand des nombres de ramification propres de K/k, et dont le dénominateur sera désigné par Δ. K/K' étant primitif et complètement ramifié de degré puissance de p, il n'a qu'un seul nombre de ramification propre v, dont le dénominateur sera désigné par Δ'. Δ et Δ' sont premiers à p.

Donc il existe un corps non-ramifié $\mathfrak{L}$ tel que $\mathfrak{L}(\pi^{1/\Delta})$ contienne tous les $\sigma\pi$, $\sigma \in V_{m-1}$. Et, de même, il existe un corps non-ramifié $\mathfrak{L}'$ tel que $\mathfrak{L}'(\pi^{1/\Delta'})$ contienne tous les $\sigma\pi$, $\sigma \in G_{K/K'}$.

Soit $\sigma_0 \in V_{m-1}$, et soit $\sigma \in G_{K/k}$. Choisissons le surcorps galoisien K^*/k de K/k de manière qu'il contienne $\mathfrak{L}(\pi^{1/\Delta})$ et $\mathfrak{L}'(\pi^{1/\Delta'})$.

$$(\sigma\sigma_0)^{(V^*)}\pi = \text{corr.}_K\,[(\text{gen.}_{K^*}\,\sigma \cap V^*)(\text{gen.}_{K^*}\,\sigma_0 \cap V^*)]\pi$$

est l'ensemble de tous les $\sigma^*\sigma_0\pi$ distincts tels que $\sigma^* \in \text{gen.}_{K^*}\,\sigma \cap V^*$. Or, puisque $K^* \supseteq \mathfrak{L}(\pi^{1/\Delta})$, on a $\sigma^*\sigma_0\pi = \text{corr.}_{\mathfrak{L}(\pi^{1/\Delta})}\,\sigma^* \cdot \sigma\pi$. Mais, puisque $\sigma^* \in V^*$, et puisque le corps $\mathfrak{L}$ qui est non-ramifié est contenu dans le corps d'inertie de K^*/k, σ^* conserve tous les éléments de $\mathfrak{L}$. D'autre part, $(\sigma^*\pi^{1/\Delta})^\Delta = \sigma^*\pi = \sigma\pi$, donc les $\sigma^*\pi^{1/\Delta}$, pour les $\sigma^* \in \text{gen.}_{K^*}\,\sigma$, ne peuvent différer que par des facteurs racines Δ-ièmes de l'unité. Donc, Δ étant premier à p, $\sigma_1^*\pi^{1/\Delta}$ et $\sigma_2^*\pi^{1/\Delta}$ (σ_1^*, $\sigma_2^* \in \text{gen.}_{K^*}\,\sigma$) ou bien coïncident, ou bien l'ordre de leur différence pour $\mathfrak{P}^{1/\Delta}$ est 1. Si, de plus, σ_1^*, $\sigma_2^* \in V^*$, l'ordre en $\mathfrak{P}^{1/\Delta}$ de $\sigma_1^*\pi^{1/\Delta} - \pi^{1/\Delta}$ et de $\sigma_2^*\pi^{1/\Delta} - \pi^{1/\Delta}$, donc aussi celui de $\sigma_1^*\pi^{1/\Delta} - \sigma_2^*\pi^{1/\Delta}$ est > 1; donc, $\sigma_1^*\pi^{1/\Delta} = \sigma_2^*\pi^{1/\Delta}$. Donc, $\text{corr.}_{\mathfrak{L}(\pi^{1/\Delta})}$ $(\text{gen.}_{K^*}\,\sigma \cap V^*)$ ne contient qu'un seul élément. Donc

$\sigma^*\sigma_0\pi$, $\sigma^* \in \text{gen.}_{K^*}\, \sigma \cap V^*$ ne prend qu'une seule valeur. Donc $(\sigma\sigma_0)^{(V^*)}$ est l'ensemble d'un seul élément.[46]

Pour des causes identiques, si $\sigma_0 \in G_{K/K'}$ et si $\sigma \in G_{K/k}$, $(\sigma\sigma_0)^{(V^*)}$ est l'ensemble d'un seul élément.[46, 47]

Soient σ et σ' deux éléments de $G_{K/K'}V_{m-1}$. Alors σ' est contenu dans un ensemble de la forme $(\sigma_1'\sigma_2')^{(V^*)}$, où $\sigma_1' \in G_{K/K'}$ et où $\sigma_2' \in V_{m-1}$. Mais cet ensemble ne contient qu'un seul élément. Donc, $(\sigma_1'\sigma_2')^{(V^*)} = \{\sigma'\}$ et $(\sigma\sigma')^{(V^*)} = (\sigma\sigma_1'\sigma_2')^{(V^*)} = ((\sigma\sigma_1')^{(V^*)}\sigma_2')^{(V^*)}$. Or, puisque $\sigma_1' \in G_{K/K'}$, $(\sigma\sigma_1')^{(V^*)}$ ne contient qu'un seul élément. Et puisque $\sigma_2' \in V_{m-1}$, il en est de même pour $((\sigma\sigma_1')^{(V^*)}\sigma_2')^{(V^*)} = (\sigma\sigma')^{(V^*)}$. Donc $(G_{K/K'}V_{m-1})^{(V^*)}$ est un hypergroupe tel que le composé de ses deux éléments quelconques est l'ensemble d'un seul élément. Donc c'est un groupe.

$V_{m-1}^{(V^*)} = V_{m-1}^{(V^*)}/V_m^{(V^*)}$ est un groupe abélien. De même, puisque K/K' est primitif, complètement ramifié et de degré puissance de p, donc $V_{0,K/K'} = G_{K/K'}$ et $V_{1,K/K'} = \{1_K\}$, on a que $G_{K/K'}^{(V^*)} = V_{0,K/K'}^{(V^*)}/V_{1,K/K'}^{(V^*)}$ est un groupe abélien. Donc, pour prouver que $(G_{K/K'}V_{m-1})^{(V^*)}$ l'est aussi, il suffit de prouver que tout $\sigma_1 \in G_{K/K'}$ est permutable avec tout $\sigma_2 \in V_{m-1}$.

Or, on a

$$\frac{(\sigma_1\sigma_2)^{(V^*)}\pi}{\pi} = \sigma'\left(\frac{\sigma_2\pi}{\pi}\right)\cdot\frac{\sigma_1\pi}{\pi},$$

où σ' est l'unique élément de l'hypergroupe de $\mathfrak{K}(\pi^{1/\Delta})$ tel que $\sigma'\pi = \sigma_1\pi$ et que $\text{gen.}_{K^*}\,\sigma' \cap V^*$ ne soit pas vide.

Mais $\dfrac{\sigma_2\pi}{\pi} - 1$ est d'ordre v_{m-1} pour $\mathfrak{P}$. Donc, il existe un nombre ρ de $\mathfrak{K}$ et un $\epsilon > 0$ tels que

$$\frac{\sigma_2\pi}{\pi} \equiv 1 + \rho(\pi^{1/\Delta})^{\Delta v_{m-1}} \qquad (\text{mod } \mathfrak{P}^{v_{m-1}+\epsilon}),$$

et que

$$\sigma'\pi^{1/\Delta} \equiv \pi^{1/\Delta} \qquad (\text{mod } \mathfrak{P}^{1/\Delta+\epsilon}).$$

Mais alors, puisque $\sigma'\rho = \rho$ et $\sigma'\mathfrak{P} = \mathfrak{P}$, on a

$$\sigma'\left(\frac{\sigma_2\pi}{\pi}\right) \equiv 1 + \rho(\pi^{1/\Delta})^{\Delta v_{m-1}} \equiv \frac{\sigma_2\pi}{\pi} \qquad (\text{mod } \mathfrak{P}^{v_{m-1}+\epsilon}),$$

[46] MM. Ore et Dresher appellent un élément c_0 d'un hypergroupe H *scalaire gauche* resp. *droit* si, pour tout $c \in H$, c_0c resp. cc_0 est l'ensemble d'un seul élément [OD, 710]. Ils démontrent que l'ensemble de scalaires gauches resp. droits de H est clos par rapport à la multiplication (qui y est univoque), quand il n'est pas vide [OD, 710]. Ils appellent $c_0 \in H$ *scalaire* s'il est à la fois scalaire gauche et scalaire droit. La notion de scalaire est due à M. Wall [W, 78]. On voit qu'en particulier l'ensemble de scalaires droits d'un hypergroupe$_D$ en est un sous-groupe s'il n'est pas vide.

[47] On voit que le groupe de scalaires droits de $V^{(V^*)}$ contient comme sous-groupe chaque groupe $G_{K/K'}^{(V^*)}$ tel que le corps K/K' soit primitif et complètement ramifié de degré puissance de p.

et

$$\frac{(\sigma_1\sigma_2)^{(V^*)}\pi}{\pi} \equiv \frac{\sigma_1\pi}{\pi}\cdot\frac{\sigma_2\pi}{\pi} \qquad (\text{mod } \mathfrak{P}^{v_m-1+\epsilon}).$$

De même,

$$\frac{(\sigma_2\sigma_1)^{(V^*)}\pi}{\pi} = \sigma'\left(\frac{\sigma_1\pi}{\pi}\right)\cdot\frac{\sigma_2\pi}{\pi},$$

où σ' est l'unique élément de l'hypergroupe de $\mathfrak{K}'(\pi^{1/\Delta'})$ tel que $\sigma'\pi = \sigma_2\pi$, et que gen.$_{K^*}$ $\sigma' \cap V^*$ ne soit pas vide. On a

$$\left(\frac{\sigma'\pi^{1/\Delta'}}{\pi^{1/\Delta'}}\right)^{\Delta'} = \frac{\sigma'\pi}{\pi} = \frac{\sigma_2\pi}{\pi}.$$

Donc, puisque Δ' est premier à p et puisque

$$\frac{\sigma'\pi^{1/\Delta'}}{\pi^{1/\Delta'}} \equiv 1 \qquad (\text{mod } \mathfrak{P}^{1/\Delta'}),$$

on a

$$\frac{\sigma'\pi^{1/\Delta'}}{\pi^{1/\Delta'}} \equiv 1 \qquad (\text{mod } \mathfrak{P}^{v_m-1})$$

et

$$\sigma'\pi^{1/\Delta'} \equiv \pi^{1/\Delta'} \qquad (\text{mod } \mathfrak{P}^{v_m-1+1/\Delta'}).$$

Or, $\dfrac{\sigma_1\pi}{\pi}$ peut se représenter comme un polynome en $\pi^{1/\Delta'}$ à coefficients entiers de $\mathfrak{K}'$. σ' conservant les nombres de $\mathfrak{K}'$, on voit que

$$\sigma'\left(\frac{\sigma_1\pi}{\pi}\right) \equiv \frac{\sigma_1\pi}{\pi} \qquad (\text{mod } \mathfrak{P}^{v_m-1+1/\Delta'}).$$

Donc, si l'on assujettit le nombre ϵ précédent d'être $\leqq 1/\Delta'$, on voit que

$$\frac{(\sigma_2\sigma_1)^{(V^*)}\pi}{\pi} \equiv \frac{\sigma_1\pi}{\pi}\cdot\frac{\sigma_2\pi}{\pi} \equiv \frac{(\sigma_1\sigma_2)^{(V^*)}\pi}{\pi} \qquad (\text{mod } \mathfrak{P}^{v_m-1+\epsilon}),$$

et que

$$\frac{(\sigma_2\sigma_1)^{(V^*)}\pi}{(\sigma_1\sigma_2)^{(V^*)}\pi} - 1 \equiv 0 \qquad (\text{mod } \mathfrak{P}^{v_m-1+\epsilon}).$$

Or, $(\sigma_1\sigma_2)^{(V^*)}\pi$ est un nombre du corps $(\sigma_1\sigma_2)^{(V^*)}K$ d'ordre 1 pour l'idéal premier $(\sigma_1\sigma_2)^{(V^*)}\mathfrak{P} = \mathfrak{P}$ de ce corps, et $(\sigma_2\sigma_1)^{(V^*)}\pi$ est un conjugué de ce nombre par rapport à $k = (\sigma_1\sigma_2)^{(V^*)}k$.

Donc

$$\omega\left(\frac{(\sigma_2\sigma_1)^{(V^*)}\pi}{(\sigma_1\sigma_2)^{(V^*)}\pi} - 1\right)$$

est égal à un des nombres de ramification de $(\sigma_1\sigma_2)^{(V^*)}K/k$. Mais ce corps étant conjugué avec K/k, il doit avoir les mêmes nombres de ramification que K/k. Donc le plus grand des nombres de ramification propres de ce corps est v_{m-1}. Donc

$$\omega\left(\frac{(\sigma_2\sigma_1)^{(V^*)}\pi}{(\sigma_1\sigma_2)^{(V^*)}\pi} - 1\right),$$

qui dépasse v_{m-1}, doit être égal à $+\infty$, c'est-à-dire $(\sigma_2\sigma_1)^{(V^*)}\pi = (\sigma_1\sigma_2)^{(V^*)}\pi$. Puisque π définit K/k, on a $(\sigma_1\sigma_2)^{(V^*)} = (\sigma_2\sigma_1)^{(V^*)}$, et tout est prouvé.

Théorème 8. *Toute suite génératrice de G_{K/K_0} (K_0 est le corps de ramification de K/k) en est une suite de composition.*

Démonstration. Soit $G_{K/K_0} = G_0 \supset \cdots \supset G_s = \{1_K\}$ une suite génératrice de G_{K/K_0}. Soit $K^{(i)}$ le corps tel que $G_{K/K^{(i)}} = G_i$. Considérons un i ($0 \leqq i < s$). Supposons qu'on ait prouvé pour un j ($i < j < s$) que G_{i+1}/G_j soit semi-invariant dans G_i/G_j, c'est-à-dire que, pour tout $\sigma \in G_i$, on ait $\sigma G_{i+1} \subseteq G_{i+1}\sigma G_j$. Or, $K^{(i)} \subset K^{(i+1)} \subseteq K^{(j)} \subset K^{(j+1)}$, et $K^{(i+1)}/K^{(i)}$ et $K^{(j+1)}/K^{(j)}$ sont primitifs. Soit $\sigma' = \text{corr.}_{K^{(j+1)}}\,\sigma$. On a, en vertu du lemme précédent, $\sigma' G_{K^{(j+1)}/K^{(j)}} \subseteq G_{K^{(j+1)}/K^{(i+1)}}\sigma$, c'est-à-dire puisque la transformation $\text{corr.}_{K^{(j+1)}}$ est un isomorphisme de $G_{K^{(j+1)}/K^{(i)}}$ avec $G_{K/K^{(i)}}/G_{K/K^{(j+1)}} = G_i/G_{j+1}$, appliquant G_{i+1}/G_{j+1} sur $G_{K^{(j+1)}}/G_{K^{(i+1)}}$ et G_j/G_{j+1} sur $G_{K^{(j+1)}}/G_{K^{(j)}}$, on a $\sigma G_j \subseteq G_{i+1}\sigma G_{j+1}$, donc $\sigma G_{i+1} \subseteq G_{i+1}\sigma G_j \subseteq G_{i+1}\cdot G_{i+1}\sigma G_{j+1} = G_{i+1}\sigma G_{j+1}$, et G_{i+1}/G_{j+1} est aussi semi-invariant dans G_i/G_{j+1}.

Comme pour $j = i + 1$, $G_{i+1}/G_j = G_{i+1}/G_{i+1}$ est bien semi-invariant dans $G_i/G_{i+1} = G_i/G_j$, on voit que $G_{i+1} = G_{i+1}/G_s$ est semi-invariant dans $G_i/G_s = G_i$, et le théorème est prouvé.

En appliquant le théorème de Jordan-Hölder on obtient la

Conséquence. *Toutes les suites génératrices d'une extension $\mathfrak{P}$-adique K/k telle que $K_0 = k$ ont la même suite d'hypergroupes (à d'ordre et à l'isomorphisme près), donc la même longueur et la même suite de degrés.*

On appelle une suite génératrice d'une extension $\mathfrak{P}$-adique K/k *régulière* si elle passe par K_0. On l'appelle *fortement régulière* si elle passe, de plus, par K_{-1}.

Théorème 9. *Toutes les suites génératrices régulières d'une extension $\mathfrak{P}$-adique K/k ont la même longueur et la même suite de degrés* (*à l'ordre près*).

Démonstration. Soient $K = K^{(0)} \supset K^{(1)} \supset K^{(2)} \supset \cdots \supset K^{(s_1)} = K_0 \supset K^{(s_1+1)} \supset \cdots \supset K^{(s_2)} = k$ et $K = K^{(0)\prime} \supset K^{(1)\prime} \supset K^{(2)\prime} \supset \cdots \supset K^{(s_1')\prime} = K_0 \supset K^{(s_1'+1)\prime} \supset \cdots \supset K^{(s_2')\prime} = k$ deux suites génératrices régulières de K/k. La partie de la première suite de $K^{(0)} = K$ jusqu'à $K^{(s_1)} = K_0$ et la partie de la deuxième suite de $K^{(0)\prime} = K$ jusqu'à $K^{(s_1')\prime} = K_0$ sont deux suites génératrices de K/K_0. Donc $s_1 = s_1'$, et ces parties de deux suites ont la même suite de degrés (à l'ordre près). D'autre part la partie de la première suite de $K^{(s_1)} = K_0$ jusqu'à $K^{(s_2)} = k$ et la partie de la deuxième suite de $K^{(s_1')\prime} = K_0$ jusqu'à $K^{(s_2')\prime} = k$ sont deux suites de composition de K_0/k. Donc chacun des corps

primitifs $K^{(i)}/K^{(i+1)}$ ($i = s_1, s_1 + 1, \cdots, s_2 - 1$) ou $K^{(i)\prime}/K^{(i+1)\prime}$ ($i = s_1'$, $s_1' + 1, \cdots, s_2' - 1$) ou bien n'est pas ramifié, ou bien est complètement ramifié de degré premier à p. Mais dans ce cas son degré doit être égal à un nombre premier. Donc la suite de degrés de ces deux suites doit être la même, à l'ordre près, et, par conséquent, $s_2 - s_1 = s_2' - s_1'$. Donc les suites génératrices initiales ont la même suite de degrés (à l'ordre près), et la même longueur.

THÉORÈME 10. *Toutes les suites génératrices fortement régulières d'une extension $\mathfrak{P}$-adique K/k ont la même suite d'hypergroupes (à l'ordre et à l'isomorphisme près).*[48]

Démonstration. Subdivisons chaque suite génératrice fortement régulière de K/k en trois suites partielles: de K à K_0, de K_0 à K_{-1} et de K_{-1} à k. La première de ces 3 parties est une suite génératrice de K/K_0 ; elle a donc la même suite d'hypergroupes (à l'isomorphisme et à l'ordre près) pour toutes les suites génératrices de K/k. La seconde partie est une suite génératrice de K_0/K_{-1} et la troisième partie en est une de K_{-1}/k. Il suffit donc de prouver, pour démontrer le théorème, que toutes les suites génératrices de K_0/K_{-1} ont la même suite d'hypergroupes (à l'ordre et à l'isomorphisme près) et qu'il en est de même pour celles de K_{-1}/k.

(a) $G_{K_0/K_{-1}} \simeq T/V$.

Or, $T^{(T^*)}/V^{(T^*)}$ est un groupe cyclique. Donc tout sous-hypergroupe de $T^{(T^*)}/V^{(T^*)}$ y est invariant. Par conséquent, tout sous-hypergroupe de T/V y est semi-invariant. Donc toute suite génératrice en est une suite de composition (et même une suite principale).

En appliquant le théorème de Jordan-Hölder pour les hypergroupes$_D$ on trouve le résultat cherché.

(b) $G_{K_{-1}/k} \simeq G_{K/k}/T$ est un groupe cyclique, d'où le résultat cherché.

Ainsi le théorème est démontré.

Les théorèmes 8, 9 et 10 peuvent se démontrer, comme je l'avais indiqué dans mon travail K2 (p. 125), à partir d'un cas particulier du théorème qui y était démontré [K2, 123–124], par l'emploi d'une formule que j'avais indiquée sans démonstration dans ma thèse [K1, 68, chapitre III, A, formule (8)] et d'un résultat de ma thèse [K1, 78]. Cette démonstration, que je publierai ailleurs, évite l'emploi du théorème de Jordan-Hölder, mais est moins directe, et la démonstration de la formule (8) du chapitre III, A de K1 exige l'emploi de méthodes peu élégantes.

Les théorèmes 9 et 10 restent encore vrais pour toutes les extensions à valuation discrète K/k telles que le "corps d'inertie régulier" de K/k au sens de M. Ostrowski [Os, 339–342] soit k/k. Il faut remplacer dans l'énoncé de ces

[48] Dans mon travail K2 (§7, p. 125) j'avais énoncé ce théorème pour toutes les suites génératrices régulières de K/k, ce qui est inexact. D'autre part, j'y avais démontré (pp. 123–124) ce théorème pour le cas particulier de suites génératrices *normales*, c'est-à-dire passant par tous les K_i ($i = -1, 0, 1, \cdots, m - 1$) de K/k et aussi (pp. 124–125) un résultat sur l'existence de suites génératrices normales satisfaisant à certaines conditions.

théorèmes K_0/k par le "corps de ramification régulier" au sens de M. Ostrowski [Os, 347–351]. Toutefois, la démonstration de ces résultats dans ce cas plus général comporte quelques complications.

11. **Extensions métagaloisiennes. Métagroupes$_D$ de la théorie de la ramification.** Une extension algébrique K/k s'appelle *métagaloisienne* s'il existe une chaîne de corps $K = K^{(0)} \supset K^{(1)} \supset K^{(2)} \supset \cdots \supset K^{(s)} = k$ telle que tous les $K^{(i-1)}/K^{(i)}$ $(i = 1, 2, \cdots, s)$ soient des extensions galoisiennes. Comme $K^{(i-1)}/K^{(i)}$ est une extension galoisienne si, et seulement si, $G_{K^{(i-1)}/K^{(i)}} \simeq G_{K/K^{(i)}}/G_{K/K^{(i-1)}}$ est un groupe [K1, 33; K2, 81], on voit que *K/k est une extension métagaloisienne si, et seulement si, $G_{K/k}$ est un métagroupe$_D$.*

Considérons l'hypergroupe $V^{(V^*)}$. Nous avons vu que, pour tout $q = 0, 1, \cdots, m-1$, $V_{q+1}^{(V^*)}$ est invariant dans $V_q^{(V^*)}$, et $V_q^{(V^*)}/V_{q+1}^{(V^*)}$ est un groupe abélien du type $(p, p, \cdots, p)$. Donc *$V^{(V^*)}$ est un métagroupe$_D$* et *l'ensemble de quotients d'une suite de composition quelconque de $V^{(V^*)}$ est formée de groupes cycliques d'ordre p.*

Or, si K_0^*/k est le corps de ramification de K^*/k, $V^{(V^*)} \simeq G_{KK_0^*/K_0^*} = V_{KK_0/K_0^*}$. Donc, puisque toute suite génératrice de KK_0^*/K_0^* est fortement régulière, toute suite génératrice de $G_{KK_0^*/K_0^*}$, donc aussi de $V^{(V^*)}$ en est une suite de composition. Donc, *l'ensemble de quotients d'une suite génératrice quelconques de $V^{(V^*)}$ est formé de groupes cycliques d'ordre p.*

Ou, encore, puisque $V^{(V^*)} = V$ quand $K_0^* = k$, cela peut se formuler ainsi:

K^*/k étant une extension $\mathfrak{P}^*$-adique galoisienne complètement ramifiée de degré puissance de p, et K/k étant une sous-extension de K^*/k, toute suite génératrice $K = K^{(0)} \supset K^{(1)} \supset K^{(2)} \supset \cdots \supset K^{(s)} = k$ de K/k est telle que, pour tout $i = 1, 2, \cdots, s$, $K^{(i-1)}/K^{(i)}$ soit une extension cyclique de degré p.

J'avais prouvé dans ma thèse [K1, 80–81] d'une autre manière, fondée sur un théorème de Sylow sur les p-groupes, le fait que $V_{K^*/K}$ appartient à une suite de composition de $V_{K^*/k}$, équivalent à l'énoncé précédent.

Considérons l'hypergroupe $T^{(T^*)}$. On avait vu déjà que $T^{(T^*)}/V^{(T^*)}$ est un groupe cyclique. On avait vu aussi que, pour tout $q = 0, 1, \cdots, m-1$, $V_{q+1}^{(T^*)}$ est semi-invariant dans $V_q^{(T^*)}$. Soit $\sigma \in V_q^{(T^*)}$. Soit δ_q le dénominateur de v_q. Soit π un nombre de K d'ordre 1 pour $\mathfrak{P}$, et soit π' un nombre tel que $(\pi')^{\delta_q} = \pi$. Désignons par $\beta_q(\sigma)$ la classe (mod $\mathfrak{P}^*$) (où $\mathfrak{P}^*$ désigne l'idéal premier d'un surcorps galoisien par rapport à k du corps K/k) à laquelle appartient $(\sigma\pi - \pi)/(\pi\pi'^{\delta_q v_q})$. J'avais prouvé [K1, 50–53; K2, 88] que *$\beta_q(\sigma)$ ne dépend que de la classe à droite suivant V_{q+1} à laquelle appartient σ et en dépend effectivement* (c'est-à-dire $\beta_q(\sigma_1) = \beta_q(\sigma_2)$ $(\sigma_1, \sigma_2 \in V_q)$ n'a lieu que si $\sigma_1 V_{q+1} = \sigma_2 V_{q+1}$), que *$M_q = \beta_q(V_q)$ est un module, admettant comme opérateurs les multiplications par toute racine δ_q-ième de l'unité et tous les isomorphismes par rapport au corps de restes de K* (mod $\mathfrak{P}^*$) (donc, si F désigne le degré absolu de l'idéal $\mathfrak{P}$ dans K, les élévations en toute puissance d'exposant p^{iF} $(i = 1, 2, \cdots, +\infty)$), et que, $\mathfrak{E}_{\delta_q}$ désignant l'ensemble de toutes les racines δ_q-ièmes de l'unité,

$V_q^{(T^*)}/V_{q+1}^{(T^*)}$ *est isomorphe à l'ensemble* M_q *organisé en hypergroupe par la loi de composition* $*$ *donnée par la formule*[49]

$$a * b = a + \mathfrak{E}_{\delta_q} b,$$

l'isomorphisme se réalisant par la correspondance $\sigma V_{q+1}/V_{q+1} \to \beta_q(\sigma)$.

J'avais prouvé dans ma thèse [K1, 78–83], d'une manière assez compliquée et non tout-à-fait intrinsèque, que

La condition nécessaire et suffisante pour que tous les nombres de ramification de K/k *soient entiers est que* $T_{K^*/K}$ *fasse partie d'une suite de composition de* $T_{K^*/k}$.

Puisque $T^{(T^*)} \simeq T_{K^*/k}/T_{K^*/K}$, on peut formuler ce résultat sous la forme du

Théorème 11. $T^{(T^*)}$ *est un métagroupe*$_D$ *si, et seulement si, tous les* v_q $(q = 0, 1, \cdots, m)$ *sont entiers.*

En voici une démonstration plus simple et intrinsèque: $V^{(T^*)}$ étant semi-invariant dans $T^{(T^*)}$, et $V_{q+1}^{(T^*)}$ l'étant, pour tout $q = 0, 1, \cdots, m - 1$, dans $V_q^{(T^*)}$, il existe une suite de composition de $T^{(T^*)}$ qui passe par tous les V_q $(q = 0, 1, \cdots, m)$.

Par conséquent, $T^{(T^*)}$ est un métagroupe$_D$ si, et seulement si, $T^{(T^*)}/V^{(T^*)}$ et, pour tout $q = 0, 1, \cdots, m - 1$, $V_q^{(T^*)}/V_{q+1}^{(T^*)}$ le sont. $T^{(T^*)}/V^{(T^*)}$ est toujours un groupe (cyclique). $V_q^{(T^*)}/V_{q+1}^{(T^*)}$ étant isomorphe à M_q organisé en hypergroupe par la loi de composition $*$, il est un métagroupe$_D$ si et seulement si ce dernier hypergroupe $(M_q, *)$ l'est. Soit v_q entier. Alors $\delta_q = 1$ et $\mathfrak{E}_{\delta_q} = 1$. Donc, si $a, b \in M_q$, on a $a * b = a + 1 \cdot b = a + b$, et l'opération $*$ est l'addition. M_q étant un module, l'hypergroupe $(M_q, *) = (M_q, +)$ est un groupe, donc un métagroupe$_D$.

Inversement, soit $(M_q, *)$ un métagroupe$_D$. Soit $M_q \supset \mu_1 \supset \mu_2 \supset \cdots \supset \mu_{s-1} \supset 1_{M_q} = \{0\}$ une suite de composition de $(M_q, *)$. Alors $(\mu_{s-1}, *) = (\mu_{s-1}, *)/(1_{M_q}, *)$ est un groupe. μ_{s-1} contient au moins un élément non nul, soit a. $(\mu_{s-1}, *)$ étant un groupe, on doit avoir $\{a\} = 0 * a = 0 + a\mathfrak{E}_{\delta_q} = a\mathfrak{E}_{\delta_q}$. Puisque $a \neq 0$, ceci exige $\mathfrak{E}_{\delta_q} = \{1\}$, donc $\delta_q = 1$, et v_q est entier. Donc $T^{(T^*)}$ est un métagroupe$_D$ si et seulement si tous les v_q $(0 \leqq q < m)$ sont entiers.

On voit, comme précédemment (parce que $(M_q, +)$ est un groupe abélien du type $(p, p, \cdots, p)$) que *si* K/k *est une sous-extension d'une extension* $\mathfrak{P}^*$*-adique galoisienne* K^*/k *telle que* $K_{-1}^* = k$, *toute suite génératrice régulière de* K/k *a la suite d'hypergroupes formée de groupes cycliques de degré premier quand les* v_q $(q = 0, 1, \cdots, m - 1)$ *sont tous entiers.*

Cherchons maintenant quelle est la condition nécessaire et suffisante pour que l'hypergroupe de Galois $G_{K/k}$ d'une extension $\mathfrak{P}$-adique K/k soit un méta-

[49] A, B étant deux ensembles d'objets susceptibles d'être additionnés (resp. multipliés) $A + B$ (resp. AB) désigne l'ensemble de tous les $a + b$ (resp. ab), $a \in A$ et $b \in B$, distincts. Par convention, si A (resp. B) est l'ensemble d'un seul élément, soit $\{a\}$ (resp. $\{b\}$), il est permis d'écrire a (resp. b) au lieu de $A = \{a\}$ (resp. $B = \{b\}$) dans $A + B$ et AB.

groupe$_D$. Désignons par Ω_a le champ de Galois de p^a éléments. Δ étant un entier rationnel premier à p, et α étant un élément d'un surchamp de Galois de Ω_a, $[\alpha]_{a,\Delta}$ va désigner l'ensemble de tous les conjugués par rapport à Ω_a des éléments de $\alpha\mathfrak{E}_\Delta$, c'est-à-dire l'ensemble de tous les éléments distincts de la forme $\epsilon\alpha^{ia}$ ($\epsilon^\Delta = 1$; $i = 0, 1, 2, \cdots, +\infty$). J'avais montré [K1, 53–54; K2, 89–91] que V_q/V_{q+1} *est isomorphe à l'ensemble* M_q *organisé en hypergroupe par la loi de composition* $\otimes$ *suivante*:

$$a \otimes b = a + [b]_{f,\delta_q} \qquad (a, b \in M_q)$$

l'isomorphisme étant réalisé par la même correspondance $\sigma V_{q+1}/V_{q+1} \to \beta_q(\sigma)$.

Considérons un ensemble fini M d'éléments d'un corps de caractéristique $p \neq 0$. Désignons par $f_M(x)$ le polynome $\prod_{\mu \in M} (x - \mu)$. M. Ore [O2, 563–564; O3, 246–247] (et, d'une autre manière, plus tard, l'auteur [K2, 116]) avait prouvé que M *est un module si et seulement si* $f_M(x)$ *est de la forme* $\sum_{i=0}^{n} \gamma_i x^{p^i}$. En particulier, si M est un module dans un surchamp Ω d'un champ de Galois Ω_a, M *admet comme opérateurs tous les isomorphismes de* Ω/Ω_a *si et seulement si tous les* γ_i ($i = 0, 1, \cdots, n$) *appartiennent à* Ω_a. Et, Δ étant un entier rationnel premier à p, et b étant l'exposant auquel appartient p (mod Δ), $M = \mathfrak{E}_\Delta M$ *si et seulement si* $\gamma_i \neq 0$ *n'a lieu que pour les* $i \equiv 0$ (mod b) [K2, 118].[50]

Désignons par $W_{a,b}$ l'anneau engendré sur le champ de Galois Ω_b par un élément z_a satisfaisant, pour tout $\alpha \in \Omega_b$, à la relation $z_a\alpha = \alpha^{p^a} z_a$, et ne satisfaisant à aucune relation qui n'est pas la conséquence de précédentes[51] (cet anneau est un cas particulier des anneaux introduits par M. Ore et appelés par lui "anneaux de polynomes non-commutatifs" [O1]). Si $\lambda, \lambda', \lambda'' \in W_{a,b}$ sont tels que $\lambda = \lambda'\lambda''$, on dira que λ' est *diviseur à gauche* et λ'' *diviseur à droite* de λ. En particulier, si λ est dans le centre de $W_{a,b}$, tout son diviseur à gauche en est aussi diviseur à droite, et inversement. On parlera dans ce cas de diviseurs de λ tout court. Un $\lambda \in W_{a,b}$ s'appelle *premier* s'il n'existe aucune décomposition $\lambda = \lambda'\lambda''$ ($\lambda', \lambda'' \in W_{a,b}$) de λ telle qu'aucun des λ', λ'' n'appartienne à Ω_b.

Soit

$$\lambda = \sum_{i=0}^{n} \gamma_i z_a^i \qquad (\gamma_i \in \Omega_b)$$

(tout $\lambda \in W_{a,b}$ peut être mis, et d'une seule manière, sous cette forme), λ peut être regardé comme opérateur dans un surchamp quelconque de Ω_b si l'on pose, pour tout élément α de ce surchamp,

$$\lambda\alpha = \sum_{i=0}^{n} \gamma_i \alpha^{p^{ai}}.$$

[50] M. Ore avait prouvé le second de ces deux résultats dans le cas particulier, où tous les γ_i ($i = 0, 1, \cdots, n$) appartiennent à Ω_b [O3, 253]. Au même endroit se trouve la démonstration du premier de ces résultats, qui est, d'ailleurs, évident.

[51] Ces anneaux avaient été introduits par M. Ore [O2, 560; O3, 244, 253]. La notation est de l'auteur [K2, 100].

Les isomorphismes par rapport à Ω_a se représentent par des puissances de z_a, et les multiplications par les éléments de Ω_b par ces mêmes éléments. D'autre part, puisque Ω_b est l'anneau engendré par $\mathfrak{E}_\Delta$, tout $\lambda \in W_{a,b}$ est un élément de l'anneau d'opérateurs engendré par les isomorphismes par rapport à Ω_a et par les multiplications par les éléments de $\mathfrak{E}_\Delta$. Donc tout module M admettant ces derniers opérateurs est un $W_{a,b}$-module et inversement. Il en résulte que *M est un $W_{a,b}$-module si, et seulement si, $f_M(x)$ peut se mettre sous la forme $\lambda_M x$, où λ_M* (appelé *quasi-polynome de M*) *est un élément de $W_{b,a}$. Inversement, si $\lambda \neq 0 \in W_{b,a}$ il existe un et un seul module M* (qui est forcément un $W_{a,b}$-module) *et un élément γ de Ω_a tels que $\gamma\lambda \cdot x = f_M(x)$* ($M = \{0\}$ si, et seulement si, $\lambda_M \in \Omega_a$). M. Ore[52] avait prouvé que *M' est un sous-$W_{a,b}$-module d'un $W_{a,b}$-module M si, et seulement si, $\lambda_{M'}$ est un diviseur à droite de λ_M dans $W_{b,a}$.* $M^{(1)}$, $M^{(2)}, \cdots, M^{(s)}$ étant des $W_{a,b}$-modules, je note [K2, 120–122] $M^{(1)} \times M^{(2)} \times \cdots \times M^{(s)}$ le $W_{a,b}$-module M tel que $\lambda_M = \lambda_{M^{(1)}}\lambda_{M^{(2)}} \cdots \lambda_{M^{(s)}}$. *Si M et $M' \subset M$ sont des $W_{a,b}$-modules, $M'' = \lambda_{M'}M$ satisfait à l'égalité $M = M'' \times M'$* [K2, 121].

(a, b) et $[a, b]$ désignant le p.p.c.m. et le p.g.c.d. des a, b, *le centre de $W_{a,b}$* (et de $W_{b,a}$) *est $W_{(a,b),[a,b]}$*; $\lambda_1, \lambda_2, \cdots, \lambda_s$ étant des éléments premiers de $W_{a,b}$, et λ' étant un élément (non nul) du centre de $W_{a,b}$, j'avais prouvé [K2, 101–102, 108–111] que *$\lambda_1\lambda_2 \cdots \lambda_s$ est diviseur d'une puissance convenable de λ' si, et seulement si, tous les λ_i ($i = 1, 2, \cdots, s$) divisent λ'.*

ψ_q désignant l'exposant auquel appartient p (mod δ_q), j'avais prouvé [K1, 54; K2, 93–94] que *$M' \subset M_q$ est un sous-hypergroupe de $(M_q, \otimes)$ si, et seulement s'il est un W_{F,ψ_q}-module*; et que si, M étant un sous-W_{F,ψ_q}-module de M_q, $M = M'' \times M'$, on a $(M, \otimes)/(M', \otimes) \simeq (M'', \otimes)$ [K2, 120–121, 124].

α étant un élément d'un surchamp du champ de Galois Ω_a, désignons par $\langle\alpha\rangle_a$ l'ensemble de tous les conjugués distincts par rapport à Ω_a de α. σ étant un élément de T, désignons par $\beta_{-1}(\sigma)$ la classe (mod $\mathfrak{P}^*$) à laquelle appartient $\dfrac{\sigma\pi}{\pi}$. On a $\beta_{-1}(\sigma_1) = \beta_{-1}(\sigma_2)$ ($\sigma_1, \sigma_2 \in T$) si et seulement si $\sigma_1 V = \sigma_2 V$. J'avais prouvé [K1, 47–48; K2, 86–87] que $h = n_{-1}:n_0$ désignant le nombre d'éléments de T/V (c'est-à-dire le plus grand facteur de e premier à p), *T/V est isomorphe à $\beta_{-1}(T) = \mathfrak{E}_h$, organisé en hypergroupe par la loi de composition $\square$ suivante*:

$$\alpha \,\square\, \beta = \alpha \cdot \langle\beta\rangle_F \qquad (\alpha, \beta \in \mathfrak{E}_h)$$

l'isomorphisme étant réalisé par la correspondance $\sigma V/V \to \beta_{-1}(\sigma)$.

Ce qui précède permet de résoudre le problème posé en démontrant le théorème suivant:

THÉORÈME 12. *$G_{K/k}$ est un métagroupe$_D$ si et seulement si les deux conditions suivantes sont satisfaites*: (1) *$h = n_{-1}:n_0 = e:n_0$ est diviseur d'une puissance de*

[52] M. Ore n'avait, d'ailleurs, donné une démonstration explicite de ce résultat que dans les cas $b = 1$ [O2, 561] et $b = a$ [O3, 247, 254]. La démonstration générale, qui n'en diffère presque pas, se trouve dans K2 (pp. 117–118).

$p^F - 1$; (2) *pour tout* $q = 0, 1, \cdots, m - 1$, v_q *est entier et* λ_{M_q} *est diviseur d'une puissance de l'élément* $z_1^F - 1 = z_F - 1$ *du centre* $W_{F,1}$ *de* $W_{1,F} = W_{\psi_q,F}$.

Démonstration. Pour la même raison que dans le théorème précédent, il faut et il suffit pour que $G_{K/k}$ soit un métagroupe$_D$ que $G_{K/k}/T$, T/V et tous les V_q/V_{q+1} ($q = 0, 1, \cdots, m - 1$) le soient. $G_{K/k}/T$ étant un groupe cyclique, l'est bien. Soit

$$G_{K_0/K_{-1}} \supset G_{K_0/K^{(1)}} \supset G_{K_0/K^{(2)}} \supset \cdots \supset G_{K_0/K^{(s-1)}} \supset G_{K_0/K_0}$$

une suite génératrice, donc une suite de composition, de $G_{K_0/K_{-1}}$. Pour que $T/V \simeq G_{K_0/K_{-1}}$ soit un métagroupe$_D$, il faut et il suffit que tous les $G_{K^{(i)}/K^{(i-1)}} \simeq G_{K_0/K^{(i-1)}}/G_{K_0/K^{(i)}}$ ($i = 1, 2, \cdots, s$; $K^{(0)} = K_{-1}$, $K^{(s)} = K_0$) soient des groupes. Or $K^{(i)}/K^{(i-1)}$ est un corps complètement ramifié d'un degré premier $q_i \neq p$, et $\prod_{i=1}^{s} q_i = h$. D'autre part, le degré absolu de l'idéal premier de $K^{(i)}$ est F. Donc $G_{K^{(i)}/K^{(i-1)}}$ est isomorphe à $(\mathfrak{E}_{q_i}, \square)$. Donc il est un groupe si et seulement si pour tout $\alpha \in \mathfrak{E}_{q_i}$, on a $\langle\alpha\rangle_F = \{\alpha\}$, c'est-à-dire $\alpha^{p^F} = \alpha$, donc $\alpha^{p^F-1} = 1$ (et dans ce cas la composition $\square$ est la multiplication ordinaire). Ceci n'a lieu que si q_i divise $p^F - 1$; donc, puisque tous les q_i divisent $p^F - 1$ si et seulement si $h = \prod_{i=1}^{s} q_i$ divise une puissance convenable de $p^F - 1$, cela est la condition nécessaire et suffisante pour que T/V soit un métagroupe$_D$.

$V_q/V_{q+1} \simeq (M_q, \otimes)$ ($q = 0, 1, \cdots, m - 1$). Soit $M_q = \mu_0 \supset \mu_1 \supset \mu_2 \supset \cdots \supset \mu_{s_q} = \{0\}$ une suite génératrice, donc une suite de composition de $(M_q, \otimes)$. Soit $\mu^{(i)}$ un W_{F,ψ_q}-module tel que $\mu_i = \mu^{(i)} \times \mu_{i+1}$ ($i = 0, 1, \cdots, s_q - 1$). On a $(\mu_i, \otimes)/(\mu_{i+1}, \otimes) \simeq (\mu^{(i)}, \otimes)$. Donc, puisqu'il n'existe aucun sous-hypergroupe $(\mu, \otimes)$ de $(M_q, \otimes)$ tel que $\mu_i \supset \mu \supset \mu_{i+1}$, il n'existe aucun sous-W_{F,ψ_q}-module μ' de $\mu^{(i)}$ autre que $\mu^{(i)}$ et $\{0\}$. Donc $\lambda_{\mu^{(i)}}$ est premier. Et on a $M_q = \mu^{(0)} \times \mu^{(1)} \times \cdots \times \mu^{(s_q-1)}$, donc $\lambda_{M_q} = \lambda_{\mu^{(0)}}\lambda_{\mu^{(1)}} \cdots \lambda_{\mu^{(s_q-1)}}$. $V_q/V_{q+1} \simeq (M_q, \otimes)$ est un métagroupe$_D$ si et seulement si tous les $(\mu^{(i)}, \otimes) \simeq (\mu_i, \otimes)/(\mu_{i+1}, \otimes)$ sont des groupes. Or, puisque $\alpha \otimes \beta = \alpha + [\beta]_{F,\delta_q}$, ceci a lieu si et seulement si pour tout $\beta \in \mu^{(i)}$, on a $[\beta]_{F,\delta_q} = \{\beta\}$ (et dans ce cas $(\mu^{(i)}, \otimes) = (\mu^{(i)}, +)$ est un groupe abélien, donc un groupe cyclique de degré premier). Puisque $[\beta]_{F,\delta_q} \supseteq \mathfrak{E}_{\delta_q}\beta$, et puisque $\mu^{(i)}$ contient des éléments non nuls, ceci exige que $\delta_q = 1$, c'est-à-dire que v_q soit entier. Ceci satisfait, on a $\psi_q = 1$ et $[\beta]_{F,\delta_q} = \langle\beta\rangle_F$. Pour que $\langle\beta\rangle_F = \{\beta\}$, il faut et il suffit que $\beta^{p^F} = \beta$, c'est-à-dire que $(z_F - 1)\cdot\beta = 0$. Donc $(\mu^{(i)}, \otimes)$ est un groupe si et seulement si v_q est entier et $\lambda_{\mu^{(i)}}$ divise $z_F - 1$ (qui est bien un élément du centre $W_{F,1}$ de $W_{\psi_q,F} = W_{1,F}$).

Mais, les $\lambda_{\mu^{(i)}}$ ($i = 0, 1, \cdots, s_q - 1$) étant tous premiers, ils divisent tous $z_F - 1$ si et seulement si $\lambda_{M_q} = \lambda_{\mu^{(0)}}\lambda_{\mu^{(1)}} \cdots \lambda_{\mu^{(s_q-1)}}$ divise une puissance convenable de $z_F - 1$, et le théorème est prouvé.

La condition que λ_{M_q} divise une puissance $(z_F - 1)^j$ de $z_F - 1$ se transcrit ainsi:

Il existe un entier j tel que, pour tout $\alpha \in M_q$, on ait

$$\alpha^{p^{Fj}} - \binom{j}{1}\alpha^{p^{F(j-1)}} + \binom{j}{2}\alpha^{p^{F(j-2)}} - \cdots + (-1)^{j-1}\binom{j}{j-1}\alpha^{p^F} + (-1)^j\alpha = 0,$$

où

$$\binom{j}{i} = \frac{j!}{i!(j-i)!}.$$

Il est à remarquer que les résultats du §9 de mon travail K2 (pp. 130–139) permettent de calculer très simplement les v_q et les λ_{M_q} ($q = 0, 1, \cdots, m - 1$) d'un corps défini par une équation d'Eisenstein à partir de coefficients de cette équation, et voir ainsi facilement si ce corps est métagaloisien ou non.

On voit, comme précédemment, que *si les conditions* (1) *et* (2) *sont satisfaites, la suite d'hypergroupes d'une suite génératrice fortement régulière quelconque de K/k est formée de groupes cycliques de degrés premiers.*

BIBLIOGRAPHY

M. KRASNER, *Sur la théorie de la ramification des idéaux des corps non-galoisiens de nombres algébriques* (Thèse, Paris), Mémoires (in 4°) de l'Académie de Belgique (Classe des Sciences), vol. 11(1937), fasc. 4, pp. 1–110. (Cité comme K1.)

M. KRASNER, *Sur la primitivité des corps $\mathfrak{P}$-adiques*, Mathematica (Cluj), vol. 13(1937), pp. 72–191. (Cité comme K2.)

O. ORE, *Theory of non-commutative polynomials*, Annals of Mathematics, (2), vol. 34(1933), pp. 480–508. (Cité comme O1.)

O. ORE, *On a special class of polynomials*, Transactions of the American Mathematical Society, vol. 35(1933), pp. 559–584. (Cité comme O2.)

O. ORE, *Contributions to the theory of finite fields*, Transactions of the American Mathematical Society, vol. 36(1934), pp. 243–274. (Cité comme O3.)

O. ORE, *Abriss einer arithmetischen Theorie galoischen Körpern*, I, Mathematische Annalen, vol. 100(1928), pp. 650–673. (Cité comme O4.)

O. ORE ET M. DRESHER, *Theory of Multigroups*, American Journal of Mathematics, vol. 60(1938), pp. 705–733. (Cité comme OD.)

A. OSTROWSKI, *Untersuchungen zur arithmetischen Theorie der Körper*, Mathematische Zeitschrift, vol. 39(1934), pp. 269–404. (Cité comme Os.)

H. S. WALL, *Hypergroups*, American Journal of Mathematics, vol. 59(1937), pp. 77–98. (Cité comme W.)

PARIS, FRANCE.

QUASI-MONOTONE TRANSFORMATIONS

By A. D. Wallace

1. **Introduction.** I suppose throughout this paper that the space A is metric, compact, connected and locally connected. Further, it is assumed that the transformation $T(A) = B$ is single valued and continuous and that the space B satisfies the conditions placed on A. The transformation T is said to be

(i) *interior* if the image of each open set is open (Stoïlow [1], G. T. Whyburn [1]);

(ii) *monotone* if the inverse of each point is a connected set (R. L. Moore [1], G. T. Whyburn [2]);

(iii) *quasi-monotone* if for any connected set H in B with a non-void interior, each component of $T^{-1}(H)$ maps onto all of H;

(iv) *light* if the inverse of each point is totally disconnected (G. T. Whyburn [1]).

My purpose is to study quasi-monotone transformations and their relation to the other types of mappings just defined and to develop certain properties of these transformations. By virtue of the well-known fact that if T is monotone the inverse of a connected set is also connected we have immediately

(1.1) Theorem. *The monotone transformation $T(A) = B$ is quasi-monotone.*

It is shown later that an interior transformation is quasi-monotone and that a light quasi-monotone mapping is interior. Moreover, the transformation $T(A) = B$ is quasi-monotone if and only if it can be factored, $T = T_2T_1$, where T_1 is monotone and T_2 is light and interior.

The most novel result of the paper is as follows: Let the mapping $T(A) = B$ be quasi-monotone, and let N_2 be the realized nerve of a one-dimensional covering of B with a finite number of continua. Then A admits a similar type of covering with a realized nerve N_1 and there exists a simplicial interior transformation $f(N_1) = N_2$. Thus, so far as one-dimensional structure is concerned, quasi-monotone and interior transformations behave in a similar fashion.

The following is an example of a quasi-monotone transformation that is neither monotone nor interior. It will serve to illustrate several of the results to be presented. Let S be a locally connected continuum and $A = S \times S$ the set of all pairs (x, y) with x and y in S. The distance function in S will be denoted by ρ. Further let D be the set of all pairs (x, x) and designate by T_1 the transformation which maps D into a point and is topological on $A - D$. Since D is manifestly homeomorphic with S, it is a continuum and the transformation just described, $T_1(A) = A'$, is monotone. Next let B be the space which results from the identification of the points $T_1(x, y)$ and $T_1(y, x)$. This

Received February 5, 1940.

second operation is an interior mapping $T_2(A') = B$ which is certainly light. The transformation $T = T_2T_1$ is easily seen to be quasi-monotone and is neither monotone nor interior. We may describe $T(A) = B$ as the single operation of replacing the metric of A by a new distance, *dist* $((x, y), (z, w))$, defined to be the smaller of the numbers $\rho(x, y)$, $\rho(y, w)$, $\rho(x, z)$, $\rho(x, w)$, $\rho(y, z)$, $\rho(y, w)$, and identifying points at a distance 0 from each other.

2. **Characterizations.** For brevity I denote by X^0 the collection of all interior points of the set X.

(2.1) THEOREM. *If $T(A) = B$ is quasi-monotone and H is a connected subset of B with $H^0 \neq 0$, then the set $T^{-1}(H)$ has only a finite number of components, each of which maps onto all of H.*

Proof. By virtue of the definition it is to be shown that $T^{-1}(H)$ has only a finite number of components. Let y be a point in H^0 contained in the open set $U \subset H$. Since B is locally connected, we may assume that U is a connected set. The set $T^{-1}(U)$ is open and so are its components. These latter sets cover the compact set $T^{-1}(y)$ and hence a finite number of them suffice to cover it. Thus $T^{-1}(U)$ has a finite number of components, and it follows that $T^{-1}(H)$ can have no more than this number since each component of $T^{-1}(H)$ intersects $T^{-1}(y)$.

(2.2) THEOREM. *In order that the transformation $T(A) = B$ be quasi-monotone each of the following conditions is both necessary and sufficient:*

(i) *If Z is a connected subset of B and $T^{-1}(Z) = Z_1 + Z_2$, where Z_1 and Z_2 are mutually separated, then $T(Z_1) = Z = T(Z_2)$.*

(ii) *If Z is a connected subset of B and S is a quasi-component of $T^{-1}(Z)$, then $T(S) = Z$.*

Proof. Assume that $T(A) = B$ is quasi-monotone and let Z be a connected subset of B, with $T^{-1}(Z) = Z_1 + Z_2$, as in (i). Assume, moreover, in denial of the conclusion, that p is a point in $Z - T(Z_1)$. It is clear, since A is compact and Z_1 is closed in $T^{-1}(Z)$, that $T(Z_1)$ is closed in Z. Then $Z - T(Z_1)$ is open in Z and hence we can find an open set U, containing the point p, such that $Z - T(Z_1) = ZU$. Let V be an open set containing p and with $\bar{V} \subset U$ and let R be the component of V which contains p. *Then we have*

(a) $p \in \bar{R}Z \subset Z - T(Z_1)$ *and*

(b) $T^{-1}(Z + R) = Z_1 + (Z_2 + T^{-1}(R))$,

where the two sets in the right member of (b) *are mutually separated as is indicated.* For we have $\bar{Z}_1Z_2 = 0$ by hypothesis; and if $\bar{Z}_1 \cdot T^{-1}(R)$ were not void, then because $T^{-1}(R)$ is open we would have $Z_1 \cdot T^{-1}(R)$ non-void and hence $T(Z_1)R \neq 0$, contrary to (a). Again, we have assumed that $Z_1\bar{Z}_2$ is empty, and if there existed points $q_n \in T^{-1}(R)$, $q_n \to q \in Z_1$, then we would have $T(q_n) \to T(q)$, $T(q_n) \in R$, so that $\bar{R}T(Z_1) \neq 0$, contrary to (a). Now $Z + R$ is a connected set with a non-vacuous interior so that each component of $T^{-1}(R + Z)$ maps onto all of $Z + R$, and hence each such component intersects $T^{-1}(p)$. How-

ever, if R_1 is the component of $T^{-1}(Z + R)$ which intersects Z_1, then by (b) it lies in Z_1. But $T^{-1}(p) \cdot Z_1 = 0$ by (a). This is a contradiction.

Suppose that (i) holds and let K be a quasi-component of $W = T^{-1}(Z)$, where Z is any connected subset of B. For each point x in $W - K$ we have a separation

$$W = P(x) + Q(x), \qquad K \subset P(x), \qquad x \in Q(x).$$

Since W is a separable metric space, there is a countable set of points $x_i \in W$ such that if $P_i = P(x_i)$ and $Q_i = Q(x_i)$ then $\sum Q_i = W - K$. Let $M_n = P_1 \cdot P_2 \cdot \cdots \cdot P_n$ and $N_n = Q_1 + Q_2 + \cdots + Q_n$. Then for each n we have a separation

$$W = M_i + N_i, \qquad M_1 \supset M_2 \supset M_3 \supset \cdots, \qquad \prod M_i = K.$$

By virtue of (i) we know that $T(M_i) = Z$, so that to complete the proof we must show that $T(\prod M_i) \supset \prod T(M_i)$. Let p be a point common to all the sets $T(M_i)$. Then there is a point z_i in $T^{-1}(p) \cdot M_i$. Since $T^{-1}(p)$ is compact, we may suppose that $z_i \to z \in T^{-1}(p) \subset W$. The sets $T^{-1}(p) \cdot M_i$ form a monotone decreasing family and hence $z_i + z_{i+1} + \cdots \subset M_i$, so that $z \in M_i$ for each integer i, and we have $z \in \prod M_i$ and hence $p = T(z) \in T(\prod M_i)$.

That (ii) implies the quasi-monotoneity of T is shown as follows: If H is a connected subset of B with H^0 non-vacuous, then H contains an open connected set R. Now $T^{-1}(R)$ is open and hence locally connected so that each component (= quasi-component) maps onto all of R. As in (2.1) there are only a finite number of these and hence $T^{-1}(H)$ has only a finite number of components. Thus each quasi-component of $T^{-1}(H)$ is a component and by assumption its transform is H.

Condition (ii) of this theorem may be regarded as a justification for the term "quasi-monotone". The following result is also related to this theorem and is in a form which is frequently useful.

(2.3) Theorem. *In order that $T(A) = B$ be quasi-monotone it is necessary and sufficient that for any region R (i.e., open connected set) in B, each component of $T^{-1}(R)$ map onto all of R.*

Proof. The necessity of the theorem being manifest, we need only state that the sufficiency follows as in the latter part of the proof of (2.2).

Corollary. *If $T(A) = B$ is quasi-monotone and Z is a continuum or a connected G_δ in B, then each component of $T^{-1}(Z)$ maps onto all of Z.*

(2.4) Theorem. *If the transformation $T(A) = B$ is interior, it is quasi-monotone.*

Proof. This is an immediate consequence of a theorem of G. T. Whyburn [1] and our result (2.2).

It is easy to construct an example (cf. §1) of a mapping that is quasi-monotone but neither interior nor monotone. With the aid of the following result we are able to show that if T is quasi-monotone and light, it is then interior.

(2.5) THEOREM. *If $T(A) = B$ is quasi-monotone and light, then for any point x in A and any neighborhood U of x there exists a connected neighborhood V of x, contained in U, such that*

(i) *$T(V)$ is open,*

(ii) *$F(T(V)) = T(F(V))$, where $F(X)$ is the boundary of X.*

Proof. Since T is light, the set $T^{-1}(y)$, $y = T(x)$, is totally disconnected and hence we can find a neighborhood W of x contained in U and such that $F(W) \cdot T^{-1}(y)$ is empty. Let R be the component of $B - T(F(W))$ which contains y and let V be the component of $T^{-1}(R)$ which contains x. Then $V \subset W \subset U$, since if this were not the case we would have $VF(W)$ non-vacuous and because $T(V) = R$ it would follow that $T(VF(W)) \subset T(V)T(F(W)) = RT(F(W))$. This is a contradiction. Since R is open, we know that $T(V) = R$ is open. Also

$$F(T(V)) = \overline{T(V)} - T(V) = T(\bar{V}) - T(V) \subset T(\bar{V} - V) = T(F(V));$$

again, because V is a component of the locally connected set $T^{-1}(R)$, we have

$$F(V) \subset F(T^{-1}(R)) \subset T^{-1}(\bar{R}) - T^{-1}(R)$$

and hence

$$T(F(V)) \subset T[T^{-1}(\bar{R}) - T^{-1}(R)] = \bar{R} - R = \overline{T(V)} - T(V) = F(T(V)).$$

This completes the proof.

Immediately from (i) we get

(2.6) THEOREM. *If $T(A) = B$ is quasi-monotone and light, it is interior and hence the conclusions of* (2.5) *hold for light interior transformations.*

In connection with another result (Wallace [1]) we get the following

COROLLARY. *In order that the interior transformation $T(A) = B$ be light it is necessary and sufficient that condition* (ii) *of* (2.5) *hold.*

3. **Product and factor theorems.** In this section we use the following notation: If $T(A) = B$ is factored, $T = T_2T_1$, we write $T_1(A) = A'$, $T_2(A') = B$. If $T_1(A) = A'$ and $T_2(A') = B$ are given, then we write $T = T_2T_1$ and $T(A) = B$.

It is readily seen that X is a component of $T^{-1}(Y) = T_1^{-1}T_2^{-1}(Y)$, $Y \subset B$, if and only if it is a component of $T_1^{-1}(Z)$, where Z is a component of $T_2^{-1}(Y)$.

(3.1) THEOREM. *If $T = T_2T_1$ and T is quasi-monotone, then so is T_2.*

Proof. Let R be a region (= open connected set) in B and R' a component of $T_2^{-1}(R)$. If K is a component of $T_1^{-1}(R')$, then K is a component of $T^{-1}(R)$ and hence

$$R = T(K) = T_2T_1(K) \subset T_2(R') \subset R.$$

Hence $T_2(R') = R$ and thus T_2 is quasi-monotone by virtue of (2.3).

(3.2) THEOREM. *If $T = T_2T_1$ where both T_1 and T_2 are quasi-monotone, then T is quasi-monotone.*

The proof of this is similar to that of (3.1).

(3.3) THEOREM. *In order that $T(A) = B$ be quasi-monotone it is necessary and sufficient that T be factorable into T_2T_1, where T_1 is monotone and T_2 is light and interior.*

Proof. If T_1 is monotone and T_2 is light and interior, then by (1.1) and (2.4) they are both quasi-monotone and hence their product is, by (3.2). If T is quasi-monotone, then it can be factored (G. T. Whyburn [2], S. Eilenberg [1]) so that T_1 is monotone and T_2 is light. But by (2.6) T_2 is quasi-monotone and hence by (3.1) it is interior.

From this result we secure at once the

(3.4) THEOREM. *A topological property of a space A is invariant under quasi-monotone transformations if and only if it is invariant under both monotone and light interior transformations.*

Since unicoherence is invariant under monotone (Kuratowski [1]) and interior (Eilenberg [2]) transformations, it follows at once that it is invariant under quasi-monotone transformations. Indeed we see more generally, by using results of Vietoris [1] and G. T. Whyburn [3], that a quasi-monotone transformation will not increase the rational first Betti number of a locally connected continuum. Other consequences of (3.4) will occur to the reader.

For the remainder of this section we assume familiarity with the cyclic structure of the space A (Kuratowski and Whyburn [1]). If M is a metric space and N is a closed subset of M, then the continuous transformation $r(M) = N$ is said to *retract* M onto N if $r(x) = x$ for each point x in N (Borsuk [1]). The transformation $T(A) = B$ is said to be *non-alternating* if for any point b in B and any separation $A - T^{-1}(b) = A_1 + A_2$ we have $A_i = T^{-1}T(A_i)$ (G. T. Whyburn [2]). The following result is to be found in the paper just cited:

LEMMA. *If S is a locally connected continuum and H is an A-set in S, then there exists a monotone retracting transformation $r(S) = H$ which is monotone on every subcontinuum of S.*

(3.5) THEOREM. *If $T(A) = B$ is quasi-monotone and H is an A-set in B, then there exists a quasi-monotone transformation $t(A) = H$, which is monotone if T is.*

Proof. Let $r(B) = H$ be the monotone transformation given by the lemma, and let $t = rT$. By (1.1) and (3.2) the transformation t is quasi-monotone.

If J is a simple closed curve in a locally connected continuum S and for every decomposition $S = H + K$ into closed sets such that HJ is an arc xuy and KJ is an arc xvy, the points x and y lie in the same component of HK, then S is said to be *unicoherent about* J (W. A. Wilson [1]).

(3.6) THEOREM. *If* (a) *E is a simple arc in S or* (b) *E is a simple closed curve about which S is not unicoherent, where S is a locally connected continuum, then there exists a quasi-monotone non-alternating retraction $T(S) = E$.*

Proof. In (a) let C be the cyclic chain $C(u, v)$, where u and v are the end-points of E, and in (b) let C be the cyclic element containing E. In either case there is a non-alternating interior retraction $f(C) = E$ (G. T. Whyburn [4]). Let $r(S) = C$ be the monotone retracting transformation of the lemma. Then if $T = fr$ we know by virtue of (1.1), (2.4) and (3.2) that T is quasi-monotone. It remains to show that T is non-alternating, and this is a consequence of a nown result (G. T. Whyburn [2]).

4. **Invariance and separation theorems.** In this section I shall be concerned with the invariance of certain very general properties under quasi-monotone and interior transformations. It will be understood that if a theorem is stated only for interior transformations there are examples to show that the result does not hold in the more general case.

(4.1) THEOREM. *If $T(A) = B$ is quasi-monotone and A is a regular curve or a rational curve, then so is B.*

Proof. Let x and y be any points in B. Then it suffices to show (K. Menger [1]) that if x and y are distinct, they can be separated in B by a set consisting respectively of a finite or a countable set. Now in each case the sets $T^{-1}(x)$ and $T^{-1}(y)$ are separated in A by a set of the required type, X. Then $T(X)$ is also of this type and $T^{-1}T(X)$ separates $T^{-1}(x)$ and $T^{-1}(y)$ in A. If $T(X)$ did not separate x and y in B, then $x + y$ would lie in a region R in $B - T(X)$. But each component of $T^{-1}(R)$ intersects both $T^{-1}(x)$ and $T^{-1}(y)$ and lies in $A - T^{-1}T(X)$. This is a contradiction.

As an immediate corollary to the proof of (4.1) we have the following, which is a necessary but not a sufficient condition.

(4.2) THEOREM. *If $T(A) = B$ is quasi-monotone, and X, Y and Z are disjoint closed sets in B, then Y separates X and Z in B if and only if $T^{-1}(Y)$ separates $T^{-1}(X)$ and $T^{-1}(Z)$ in A.*

Let P be a topological property of closed sets. Then P is *additive* if whenever the closed sets X and Y have P their sum has P; *hereditary*, if whenever the closed set X has P every closed subset of X also has P. A locally connected continuum has property $F(P)$ if the boundary of every one of its regions has property P.

(4.3) THEOREM. *If the property P is invariant under continuous transformations and A has property $F(P)$ and $T(A) = B$ is quasi-monotone, then B has property $F(P)$.*

Proof. Let R be a region in B and S a component of $T^{-1}(R)$. Then, as in the proof of (2.5), it may be seen that $T(F(S)) = F(T(S)) = F(R)$. But $F(S)$ has P and P is invariant under T, so that $F(R)$ has P.

As a corollary we have the following (Harrold [1]).

Corollary. *The property of a locally connected continuum being a hereditary arc-sum is invariant under monotone and interior transformations.*

(4.4) Theorem. *If $T(A) = B$ is interior and A has property $F(P)$, where P is additive, then for each y in B the set $T^{-1}(y)$ has property P.*

Proof. Let R be a component of $B - y$ so that $F(R) = y$. Then (Wallace [1]) $T^{-1}(y) = T^{-1}(F(R)) = F(T^{-1}(R))$. But $T^{-1}(R)$ is the sum of a finite number of components, $R_1, R_2, \cdots, R_n$, and hence $T^{-1}(y) = F(R_1) + F(R_2) + \cdots + F(R_n)$ and hence has property P.

Corollary. *If $T(A) = B$ is interior and all but a countable number of the points of A are local separating points, then each set $T^{-1}(y)$ is countable.*

A locally connected continuum will be said to have property $N(P)$ if each point is contained in an arbitrarily small region whose boundary has property P.

(4.5) *Let $T(A) = B$ be light and let P be a hereditary property invariant under T^{-1}. If B has property $N(P)$, so also has A.*

Proof. Let U be a neighborhood of the point x in A. We may suppose that $F(U)$ does not intersect the set $T^{-1}(y)$, where $y = T(x)$. Let R be the component of $B - T(F(U))$ which contains the point y and denote by V the component of $T^{-1}(R)$ which contains x. Then V is a subset of U. There is a neighborhood W of y which is a subset of R and such that $F(W)$ has property P. If S is the component of W which contains y, then $F(S) \subset F(W)$ and $F(S)$ has property P. Let Q be the component of $T^{-1}(S)$ which contains x. It follows that $Q \subset V \subset U$. Now

$$F(Q) \subset F[T^{-1}(S)] = \overline{T^{-1}(S)} - T^{-1}(S) \subset T^{-1}(F(S)),$$

so that $F(Q)$ has property P.

From this we get the following corollaries (W. T. Puckett [1]):

Corollary. *If T is finite and B is a regular curve, then so is A. If T is countable and B is a rational curve, then so is A.*

The following result is much stronger than (4.2) but only because T is restricted to be interior.

(4.6) Theorem. *If $T(A) = B$ is interior, then the closed set X separates B irreducibly between the points p and q if and only if the set $T^{-1}(X)$ separates A irreducibly between the sets $T^{-1}(p)$ and $T^{-1}(q)$.*

Proof. If X separates B irreducibly between p and q, then (G. T. Whyburn [5]) $X = F(G_m)$, where G_m is the component of $B - X$ containing the point m, $m = p$ or $m = q$. Hence

$$T^{-1}(X) = T^{-1}(F(G_m)) = F(T^{-1}(G_m))$$

and $T^{-1}(G_m)$ is the sum of a finite number of components each of which maps onto all of G_m. Also $T^{-1}(G_m)$ is the sum of a finite number of components each

of which maps onto G_m and these are exactly those components of $A - T^{-1}(X)$ which intersect $T^{-1}(m)$. Consequently $T^{-1}(X)$ separates A irreducibly between $T^{-1}(p)$ and $T^{-1}(q)$.

If $T^{-1}(X)$ separates A irreducibly between the set $T^{-1}(p)$ and $T^{-1}(q)$, then by virtue of (4.2) X separates B between p and q. Let C be a subset of X which separates B irreducibly between p and q. Then $T^{-1}(C)$ irreducibly separates A between $T^{-1}(p)$ and $T^{-1}(q)$ and is contained in $T^{-1}(X)$. Accordingly we have $T^{-1}(C) = T^{-1}(X)$ and thus $C = X$.

5. **One-dimensional structure.** By a *covering* C: $S = \sum S_i$ of a space S we mean a decomposition of S into a finite number of closed sets $S_i \neq 0$; by the *dimension of* C is meant the greatest integer k such that there are $k + 1$ sets of C with a non-void intersection. A covering C is said to be *monotone* if the sets S_i are connected. With a covering C of S we may associate an abstract complex, $N(C)$, called the *nerve of* C, whereby each S_i is the correspondent of one and only one vertex of $N(C)$ and the vertices corresponding to the sets S_{i_1}, $S_{i_2}, \cdots, S_{i_n}$ are spanned by an $(n - 1)$-simplex if and only if the sets have a non-null product. Since we are concerned only with one-dimensional coverings, the nerve of such a covering can be "realized" in Euclidean 3-space as a linear graph so subdivided that each edge is uniquely determined by its ends (Alexandroff-Hopf [1]).

A continuous transformation $f(K_1) = K_2$ of the linear graph K_1 onto the linear graph K_2 is said to be *simplicial* if K_1 and K_2 are so subdivided as to be 1-complexes, and f maps vertices and edges of K_1 onto vertices and edges of K_2 and is topological on each edge. With these preliminaries we may state the following

(5.1) THEOREM. *Let $C_2 : B = \sum B_i$ be a one-dimensional monotone covering and let $T(A) = B$ be quasi-monotone. There exists a one-dimensional covering C_1 : $A = \sum A_i$ and a simplicial interior transformation $f(N_1) = N_2$ of the realized nerve N_1 of C_1 onto the realized nerve N_2 of C_2.*

Proof. Let s be less than one-half the Lebesgue number of the covering C_2 (Alexandroff-Hopf [1]). Let B_i' be the closure of the component of the s-neighborhood of B_i which contains B_i. To save notation we assume that $B_i = B_i'$. For each i let $T^{-1}(B_i) = A_{i1} + \cdots + A_{ik_i}$, where A_{ij} is a component of $T^{-1}(B_i)$, and let C_1 be the covering $A = \sum A_{ij}$. Let a_{ij} and b_i be the vertices of the realized nerves N_1 and N_2 corresponding to the sets A_{ij} and B_i. Since for each i and j we have $T(A_{ij}) = B_i$, we set $f(a_{ij}) = b_i$. If $a_{i_1j_1}$ and $a_{i_2j_2}$ are spanned by a 1-simplex of N_1, then b_{i_1} and b_{i_2} are spanned by a 1-simplex of N_2 since if $A_{i_1j_1}$ and $A_{i_2j_2}$ intersect then so do B_{i_1} and B_{i_2}. We take f to be a topological mapping of the edge $(a_{i_1j_1}a_{i_2j_2})$ onto the edge $(b_{i_1}b_{i_2})$. If b_{i_1} and b_{i_2} are spanned by a 1-simplex, then B_{i_1} and B_{i_2} intersect. Hence there will be numbers j_1 and j_2 so that $A_{i_1j_1}$ and $A_{i_2j_2}$ intersect and there is a corresponding simplex spanning $a_{i_1j_1}$ and $a_{i_2j_2}$. Therefore f is well-defined, continuous, simplicial and maps N_1 onto N_2. It remains to show that f is interior.

If U is an open subset of N_1 and y is a point of $f(U)$, then if y is not a vertex of N_2, no point of $f^{-1}(y)$ is a vertex of N_1. Hence if x is a point of U and $f^{-1}(y)$ as well, we can find an open arc E containing x, lying interior to a 1-simplex of N_1, contained in U and which is a neighborhood of x. Then $f(E)$ is interior to a 1-simplex of N_2, is an open arc and lies in $f(U)$. Thus each non-vertex in $f(U)$ is an interior point of $f(U)$. Consider now the case in which y is a point of $f(U)$ which is a vertex of N_1. Then each point of $f^{-1}(y)$ is a vertex. Let x be any point of $f^{-1}(y) \cdot U$. We may suppose the notation so chosen that $b_1 = y$ and $a_{11} = x$, and further so that $a_{21}, a_{31}, \cdots, a_{n1}$ are the vertices of N_1 adjacent to a_{11}. Then $A_{11} \cdot A_{j1} \neq 0$ and hence $B_1 \cdot B_j \neq 0$, $j = 2, 3, \cdots, n$. Thus each edge on a_{11} maps onto an edge on b_1, $b_1b_k = f(a_{11}a_{k1})$. Moreover, if b_j is the vertex of an edge of N_2 on b_1, then $B_1 \cdot B_j \neq 0$; it is readily seen that this implies that $A_{11} \cdot T^{-1}(B_j) \neq 0$ so that A_{11} intersects a set A_{j1} and hence the edge b_1b_j is the image under f of the edge $a_{11}a_{j1}$. We have shown that the set of all edges on b_1 is covered by the image of the set of all edges on a_{11}. Now since a neighborhood of x is a set of arcs open at one end and pairwise disjoint except for x, a neighborhood V of x, contained in U, will map into a set of arcs open at one end and pairwise disjoint except for y. This set of arcs, in virtue of our previous remarks, will form a neighborhood of y and will lie in $f(U)$. This completes the proof.

Remark. It is clear that throughout this section we could have assumed the elements in the covering to be open rather than closed sets.

COROLLARY. *If A is unicoherent and $T(A) = B$ is quasi-monotone, then B is unicoherent.*

Proof. In order that a locally connected continuum be unicoherent it is necessary and sufficient that the nerve of every one-dimensional monotone covering be acyclic.[1] If C_2 is any one-dimensional covering of B, then its nerve N_2 is the interior image of the nerve N_1 of a similar covering of A. But N_1 is acyclic and hence so is N_2 (G. T. Whyburn [1]).

BIBLIOGRAPHY

ALEXANDROFF-HOPF.
1. *Topologie*, I, Berlin, 1935.

K. BORSUK.
1. Fundamenta Mathematicae, vol. 17(1931), p. 161.

S. EILENBERG.
1. Fundamenta Mathematicae, vol. 22(1934), p. 292.
2. Fundamenta Mathematicae, vol. 24(1935), p. 175.
3. Fundamenta Mathematicae, vol. 27(1936), p. 153.
4. Fundamenta Mathematicae, vol. 29(1937), p. 101.

[1] For this condition see Eilenberg [3] and [4] and Hopf [1]. I wish to take this opportunity to thank Dr. Eilenberg for kindly pointing out to me that Theorems 1 and 5 of my paper (Wallace [4]) are immediate consequences of his results in the papers just cited. I regret that adequate reference was not given to Dr. Eilenberg's work.

O. G. Harrold.
1. This Journal, vol. 5(1939), p. 111.
H. Hopf.
1. Fundamenta Mathematicae, vol. 28(1937), p. 33.
C. Kuratowski.
1. Fundamenta Mathematicae, vol. 11(1928), p. 172.
C. Kuratowski and G. T. Whyburn.
1. Fundamenta Mathematicae, vol. 16(1930), p. 305.
K. Menger.
1. *Kurventheorie*, Berlin, 1932.
R. L. Moore.
1. *Foundations of Point Set Theory*, American Mathematical Society Colloquium Publications, vol. 13, New York, 1932.
W. T. Puckett.
1. American Journal of Mathematics, vol. 61 (1939), p. 750.
S. Stoïlow.
1. *Leçons sur les Principes Topologiques de la Théorie des Fonctions*, Paris, 1939.
L. Vietoris.
1. Mathematische Annalen, vol. 97(1927), p. 454.
A. D. Wallace.
1. American Journal of Mathematics, vol. 61(1939), p. 757.
2. This Journal, vol. 6(1939), p. 31.
G. T. Whyburn.
1. This Journal, vol. 3(1937), p. 370.
2. American Journal of Mathematics, vol. 56(1934), p. 294.
3. This Journal, vol. 4(1938), p. 1.
4. Annals of Mathematics, vol. 40(1939), p. 914.
5. Fundamenta Mathematicae, vol. 13(1927), p. 42.
W. A. Wilson.
1. American Journal of Mathematics, vol. 55(1933), p. 135.

The University of Virginia.

A CLASS OF FUNCTIONS BOUNDED IN THE UNIT CIRCLE

By H. S. Wall

1. **Introduction.** The principal object of this note is to show how continued fractions may be used to obtain inequalities for certain functions, particularly, for *moment generating functions* of the form

$$f(x) = \int_0^1 \frac{d\phi(u)}{1 + xu}, \tag{1.1}$$

where $\phi(u)$ is bounded and monotone non-decreasing in the interval $0 \leqq u \leqq 1$. To do this, we shall make use of results contained in three recent papers ([1], [3], [4]).[1]

2. **Some remarks on continued fractions.** It is easy to see that if $g_n \neq 1$ $(n = 1, 2, 3, \cdots)$, then the continued fraction

$$1/1 - g_1/1 - (1 - g_1)g_2/1 - (1 - g_2)g_3/1 - \cdots \tag{2.1}$$

and the series

$$1 + \sum_{i=1}^{\infty} \frac{g_1 g_2 \cdots g_i}{(1 - g_1)(1 - g_2) \cdots (1 - g_i)} \tag{2.2}$$

are equivalent. In fact, if A_n/B_n is the n-th approximant of the continued fraction, then $A_n = (1 - g_1)(1 - g_2) \cdots (1 - g_{n-1})S_n$, $B_n = (1 - g_1)(1 - g_2) \cdots (1 - g_{n-1})$, where S_n is the sum of the first n terms of (2.2). Consequently, if the g_n's are real and $0 \leqq g_n < 1$ $(n = 1, 2, 3, \cdots)$, the reciprocal of (2.1) is necessarily convergent, and hence the continued fraction

$$g_1/1 + (1 - g_1)g_2x_1/1 + (1 - g_2)g_3x_2/1 + \cdots \tag{2.3}$$

converges if $x_1 = x_2 = x_3 = \cdots = -1$ and is equal, for these special values of the x_n's, to

$$1 - \left\{1 + \sum_{i=1}^{\infty} \frac{g_1 g_2 \cdots g_i}{(1 - g_1)(1 - g_2) \cdots (1 - g_i)}\right\}^{-1}. \tag{2.4}$$

Moreover ([3], pp. 159–160), if $0 < g_1 < 1$ the continued fraction (2.3) converges uniformly over the domain

$$D: \ |x_n| \leqq 1 \qquad (n = 1, 2, 3, \cdots);$$

Received February 20, 1940; presented to the American Mathematical Society (under slightly different title), April 26, 1940.

[1] The numbers in brackets refer to the bibliography.

and if $f(x)$, $x = (x_1, x_2, x_3, \cdots)$, is the function represented, then

$$M(f) = \underset{x \in D}{\text{l.u.b.}} |f(x)| \leqq 1, \tag{2.5}$$

and this upper bound is given by (2.4), which is the value of $f(x)$ when $x = (-1, -1, -1, \cdots)$. In case some g_n is 0, the continued fraction is equal to the rational fraction obtained by stopping with the first identically vanishing partial quotient.

We shall include in our discussion terminating continued fractions of the form

$$f(x) = g_1/1 + (1 - g_1)g_2x_1/1 + \cdots + (1 - g_{k-1})g_kx_{k-1}/1 + (1 - g_k)x_k/1, \tag{2.6}$$

where $0 < g_n < 1$ $(n = 1, 2, 3, \cdots, k)$. It is easily seen that we can determine numbers C, h_n such that

$$C > 1, \qquad 0 < h_n < 1 \qquad (n = 1, 2, 3, \cdots, k + 1; h_n = 0 \text{ if } n > k + 1)'$$

and such that (2.6) is identical with

$$C \cdot \{h_1/1 + (1 - h_1)h_2x_1/1 + (1 - h_2)h_3x_2/1 + \cdots\}$$

if $x \in D$. Then, by the preceding results, we see that

$$M(f) = C\left[1 - \left\{1 + \sum_{i=1}^{\infty} \frac{h_1h_2 \cdots h_i}{(1 - h_1)(1 - h_2) \cdots (1 - h_i)}\right\}^{-1}\right]. \tag{2.7}$$

Now if we allow h_{k+1} to approach 1, it is clear that the right member of (2.7) has the limit 1, so that $M(f) = 1$.

We may summarize these remarks as

THEOREM 2.1. *If $g_1, g_2, g_3, \cdots$ are real numbers and $0 \leqq g_n \leqq 1$ $(n = 1, 2, 3, \cdots)$, then the continued fraction*

$$g_1/1 + (1 - g_1)g_2x_1/1 + (1 - g_2)g_3x_2/1 + \cdots$$

represents a function $f(x)$, $x = (x_1, x_2, x_3, \cdots)$, for which $M(f) \leqq 1$, provided we agree that in case some g_n is 0 or 1 the continued fraction shall terminate with the first identically vanishing partial quotient. In any event, $M(f)$ is given by the expression (2.4) if we understand that the value of (2.4) is to be 1 in case a term of the series appearing there is infinite. Moreover,

$$M(f) = f(-1, -1, -1, \cdots).$$

We shall now give a better estimate for the functional values $f(x)$ when x is in D. To indicate the dependence of $M(f)$ upon the g_n's we shall write $M(f) = G(g_1, g_2, g_3, \cdots)$.

THEOREM 2.2. *If $f(x)$ is the function defined in Theorem 2.1 and $p = G(g_2, g_3, g_4, \cdots)$, then if $x \in D$ we have the inequality*

$$\left|f(x) - \frac{g_1}{1 - p^2(1 - g_1)^2}\right| \leqq \frac{pg_1(1 - g_1)}{1 - p^2(1 - g_1)^2}. \tag{2.8}$$

The theorem is obviously true if $g_1 = 0$ or 1. In every other case, put $f_1(x) = g_2/1 + (1 - g_2)g_3x_2/1 + (1 - g_3)g_4x_3/1 + \cdots$. Then, $M(f_1) = p$, and therefore

$$|x_1f_1(x)| = \left|\frac{g_1 - f(x)}{(1 - g_1)f(x)}\right| \leqq p, \qquad \text{if } x \in D.$$

From this inequality the theorem readily follows.

If p is replaced by its maximum possible value, $p = 1$, (2.8) takes the simpler form:

$$\left|f(x) - \frac{1}{2 - g_1}\right| \leqq \frac{1 - g_1}{2 - g_1}, \qquad \text{if } x \in D; \tag{2.9}$$

and, *a fortiori*,[2]

$$|f(x) - \tfrac{1}{2}| \leqq \tfrac{1}{2}, \quad \text{if } x \in D.$$

From Theorem 2.2 we see that

$$\Re(f) \geqq \frac{g_1}{1 + p(1 - g_1)}, \qquad \text{if } x \in D,$$

where $\Re(f)$ denotes the real part of $f(x)$. Consequently, if $f(x)$ vanishes for any x in D, it must vanish identically in D.

3. **Moment generating functions.** A large number of the familiar examples of continued fraction expansions for special functions have the form (2.3) where $0 < g_n < 1$, $x_n = e_nx$, $|e_n| \leqq 1$ $(n = 1, 2, 3, \cdots)$. One may readily verify, for example, that any continued fraction

$$1/1 + b_2x/1 + b_3x/1 + b_4x/1 + \cdots$$

can be put in the form

$$1/1 + (1 - g_1)g_2e_1x/1 + (1 - g_2)g_3e_2x/1 + \cdots,$$

where $0 < g_n < 1$, $|e_n| \leqq 1$ $(n = 1, 2, 3, \cdots)$, if

$$b_n \neq 0, \qquad |b_2| < \tfrac{1}{2}, \qquad |b_{2n-1}| + |b_{2n}| \leqq \tfrac{1}{2} \qquad (n = 2, 3, 4, \cdots). \tag{3.1}$$

In fact, it suffices to take $g_{2n} = \frac{1}{2}$, $g_{2n-1} = 1 - 2|b_{2n}|$ $(n = 1, 2, 3, \cdots)$. Some cases where the equations

$$a_1 = g_1, \qquad a_n = (1 - g_{n-1})g_n \qquad (n = 2, 3, 4, \cdots)$$

can be satisfied with $0 < g_n < 1$ are

(a) $0 < a_1 \leqq \frac{1}{2}$, $0 < a_n \leqq \frac{1}{4}$ $(n = 2, 3, 4, \cdots)$;

(b) $a_1 < \frac{1}{2}$, $a_n > 0$, $a_{2n-1} + a_{2n} \leqq \frac{1}{2}$ $(n = 1, 2, 3, \cdots)$.

[2] If w satisfies the inequality $|w - \frac{1}{2}| < \frac{1}{2}$, and g is chosen sufficiently near 1, the special function $f(x) = (1 - g)/(1 + gx)$, $0 < g < 1$, is in E and takes on the value w for some x, $|x| \leqq 1$. Hence this inequality cannot be improved for the class E.

As an illustration, consider the function ([2], p. 353)

$$e^x - \left\{1 + \frac{x}{1!} + \frac{x^2}{2!} + \cdots + \frac{x^{n-1}}{(n-1)!}\right\} = \frac{x^n}{n!}\left\{\frac{1}{1} - \frac{b_2 x}{1} + \frac{b_3 x}{1} - \cdots\right\},$$

where, for $k = 1, 2, 3, \cdots$, $b_{2k} = (n + k - 1)/(n + 2k - 2)(n + 2k - 1)$, $b_{2k+1} = k/(n + 2k - 1)(n + 2k)$. Here the conditions (3.1) hold, and we find that $g_1 = (n - 1)/(n + 1)$ $(n \geqq 2)$. On applying (2.9) to this function there results the inequality

$$\left| \frac{e^x - T_n}{\left(\dfrac{x^{n+1}}{(n+1)!}\right)} - \frac{(n+2)^2}{n(n+4)} \right| \leqq \frac{2(n+2)}{n(n+4)}, \qquad \text{if } |x| \leqq 1,\ n \geqq 1,$$

where T_n denotes the sum of the first $n + 1$ terms of the exponential series.

Every moment generating function (1.1) has a continued fraction representation of the form

$$f(x) = h_0/1 + h_1 x/1 + (1 - h_1)h_2 x/1 + (1 - h_2)h_3 x/1 + \cdots, \tag{3.2}$$

in which $0 \leqq h_n \leqq 1$ $(n = 1, 2, 3, \cdots)$, $h_0 \geqq 0$, with the established convention regarding termination of the continued fraction if some h_n is 0 or 1. Conversely, every such continued fraction represents a moment generating function [4]. Moreover, $M(f) \leqq 1$ if and only if (3.2) can be thrown into the form

$$f(x) = g_1/1 + (1 - g_1)g_2 x/1 + (1 - g_2)g_3 x/1 + \cdots, \tag{3.3}$$

where $0 \leqq g_n \leqq 1$ $(n = 1, 2, 3, \cdots)$. If in (3.2), $h_0 > 0$, $0 \leqq h_n < 1$ $(n = 1, 2, 3, \cdots)$, then $M(f) < +\infty$ if and only if the series

$$1 + \sum_{i=1}^{\infty} \frac{h_1 h_2 \cdots h_i}{(1 - h_1)(1 - h_2) \cdots (1 - h_i)} \tag{3.4}$$

is convergent; and if the sum of the latter series is S, then $M(f) = h_0 S$. Again, if $c_0 - c_1 x + c_2 x^2 - \cdots$ is the power series for $f(x)$, then $M(f) = c_0 + c_1 + c_2 + \cdots$. The proofs of these statements will be found in [4].

An important example is the hypergeometric function

$$F(\alpha, 1, \gamma; -x) = 1 - \frac{\alpha}{\gamma} x + \frac{\alpha(\alpha + 1)}{\gamma(\gamma + 1)} x^2 - \cdots,$$

where α, γ are real and $0 < \alpha < \gamma$. The corresponding continued fraction is of the form (3.2) with ([2], p. 348)

$$\begin{aligned} h_0 &= 1, \qquad h_{2n} = n/(\gamma + 2n - 1), \\ h_{2n-1} &= (\alpha + n - 1)/(\gamma + 2n - 2) \end{aligned} \qquad (n = 1, 2, 3, \cdots). \tag{3.5}$$

By means of the Gauss test we find that (3.4) converges in this case if and only if $\gamma - \alpha > 1$.

Suppose that this condition is satisfied, and denote by H the sum of the series (3.4) with the h_n's from (3.5). Then the modulus of the moment gen-

erating function $F(\alpha, 1, \gamma; -x)/H$ is $\leqq 1$ if $|x| \leqq 1$. Hence the inequality (2.9) gives

THEOREM 3.1. *If α, γ are real and positive and $\gamma - \alpha > 1$, then*

$$\left| F(\alpha, 1, \gamma; -x) - \frac{H^2}{2H-1} \right| \leqq \frac{H(H-1)}{2H-1}, \qquad \text{if } |x| \leqq 1, \tag{3.6}$$

where

$$\begin{aligned} H &= F(\alpha, 1, \gamma; 1) \\ &= 1 + \sum_{n=0}^{\infty} \frac{\alpha(\alpha+1)\cdots(\alpha+n)}{\gamma(\gamma+1)\cdots(\gamma+n)} \\ &= 1 + \sum_{n=0}^{\infty} \frac{\alpha(\alpha+1)\cdots(\alpha+n)n!(\gamma+2n+1)}{\gamma(\gamma+1)\cdots(\gamma+n)(\gamma-\alpha)(\gamma-\alpha+1)\cdots(\gamma-\alpha+n)}. \end{aligned}$$

If $x = -1$, equality holds in (3.6); *and if $|x| \leqq 1$, the real part of $F(\alpha, 1, \gamma; -x)$ is not less than $H/(2H-1)$.*

4. **Transformation of moment generating functions.** Let $f(x)$ be a moment generating function with continued fraction (3.3) so that $M(f) \leqq 1$. We shall denote by E the class of all such functions. A comparison of (3.2), (3.3) reveals that an arbitrary moment generating function $f_1(x)$ can be written in the form

$$f_1(x) = f_1(0)/[1 + xf(x)], \tag{4.1}$$

where $f(x) \in E$.

There is one simple way in which one may transform the continued fraction (3.3) so that the new continued fraction will represent a function of E, namely, by replacing g_n by $1 - g_n$ for one or more values of n. It is easily seen that a finite number of these substitutions effects upon the function a linear transformation with polynomial coefficients. When the substitution is made for infinitely many values of n, the resulting transformation of $f(x)$ may or may not be of this linear character. In particular, we have this theorem:

THEOREM 4.1. *If*

$$\begin{gathered} f(x) = g_1/1 + (1-g_1)g_2x/1 + (1-g_2)g_3x/1 + (1-g_3)g_4x/1 + \cdots \\ (0 \leqq g_n \leqq 1,\ n = 1, 2, 3, \cdots), \end{gathered} \tag{4.2}$$

then

$$\begin{aligned} \frac{1-f(x)}{1+xf(x)} &= (1-g_1)/1 + g_1(1-g_2)x/1 \\ &\quad + g_2(1-g_3)x/1 + g_3(1-g_4)x/1 + \cdots; \end{aligned} \tag{4.3}$$

$$\begin{aligned} 1 - f\left(\frac{-x}{1+x}\right) &= (1-g_1)/1 + g_1g_2x/1 \\ &\quad + (1-g_2)(1-g_3)x/1 + g_3g_4x/1 + \cdots; \end{aligned} \tag{4.4}$$

$$\frac{f\left(\frac{-x}{1+x}\right)}{1+x-xf\left(\frac{-x}{1+x}\right)} = g_1/1 + (1-g_1)(1-g_2)x/1 + g_2g_3x/1 + (1-g_3)(1-g_4)x/1 + \cdots. \tag{4.5}$$

The continued fractions in (4.3), (4.4), (4.5) *are obtained from that in* (4.2) *by replacing* g_n *by* $1 - g_n$, g_{2n-1} *by* $1 - g_{2n-1}$, g_{2n} *by* $1 - g_{2n}$ $(n = 1, 2, 3, \cdots)$, *respectively.*

The relation (4.3) results from the continued fraction identity ([4], p. 166):

$$\left[1 + \cfrac{g_1x}{1 + \cfrac{(1-g_1)g_2x}{1 + \cfrac{(1-g_2)g_3x}{1+\cdots}}}\right] \cdot \left[1 + \cfrac{(1-g_1)x}{1 + \cfrac{g_1(1-g_2)x}{1 + \cfrac{g_2(1-g_3)x}{1+\cdots}}}\right] \equiv 1 + x; \tag{4.6}$$

(4.4) follows from a theorem of Garabedian and Wall ([1], p. 191); and (4.5) is simply the result of applying (4.4) to (4.3).

In connection with (4.6) it should be remarked that the identity maintains in the sense that if the continued fractions are replaced by their corresponding power series, then the result is a formal power series identity. *It will be seen that the proof of this identity, given in* [4], *holds if the* g_n*'s are arbitrary formal power series in* x.

5. **Inequalities for functions of** E. If $f(x) \in E$, we may write down at once the inequality (2.8) for $f(x)$ and for functions such as (4.3), (4.4), (4.5) of Theorem 4.1. Suppose (4.2) holds and put

$$\begin{aligned} p &= G(g_2, g_3, g_4, g_5, \cdots), \\ q &= G(1-g_2, 1-g_3, 1-g_4, 1-g_5, \cdots), \\ r &= G(g_2, 1-g_3, g_4, 1-g_5, \cdots), \\ s &= G(1-g_2, g_3, 1-g_4, g_5, \cdots), \end{aligned} \tag{5.1}$$

where the notation is that of §2. Noting that as x ranges over the domain $|x| \leq 1$, $-x/(1+x)$ ranges over the half-plane $\Re\left(\frac{-x}{1+x}\right) \geq -\frac{1}{2}$, we then have the theorem which follows:

THEOREM 5.1. *If* $f(x)$ *is a function of* E *given by* (4.2), *then*

$$\left| f(x) - \frac{g_1}{1 - p^2(1-g_1)^2} \right| \leq \frac{pg_1(1-g_1)}{1-p^2(1-g_1)^2}, \qquad |x| \leq 1; \tag{5.2}$$

$$\left| \frac{1-f(x)}{1+xf(x)} - \frac{1-g_1}{1-q^2g_1^2} \right| \leq \frac{qg_1(1-g_1)}{1-q^2g_1^2}, \qquad |x| \leq 1; \tag{5.3}$$

$$\left| f(x) - \frac{g_1(1-r^2g_1)}{1-r^2g_1^2} \right| \leq \frac{rg_1(1-g_1)}{1-r^2g_1^2}, \qquad \Re(x) \geq -\tfrac{1}{2}; \tag{5.4}$$

$$(5.5)\qquad \left|\frac{(1+x)f(x)}{1+xf(x)} - \frac{g_1}{1-s^2(1-g_1)^2}\right| \leqq \frac{sg_1(1-g_1)}{1-s^2(1-g_1)^2}, \quad \Re(x) \geqq -\tfrac{1}{2}.$$

In these inequalities p, q, r, s are given by (5.1). *They may be replaced by* 1, *in which event the inequalities assume the simple forms*

$$(5.6)\qquad \left|f(x) - \frac{1}{2-g_1}\right| \leqq \frac{1-g_1}{2-g_1}, \qquad |x| \leqq 1;$$

$$(5.7)\qquad \left|\frac{1-f(x)}{1+xf(x)} - \frac{1}{1+g_1}\right| \leqq \frac{g_1}{1+g_1}, \qquad |x| \leqq 1;$$

$$(5.8)\qquad \left|f(x) - \frac{g_1}{1+g_1}\right| \leqq \frac{g_1}{1+g_1}, \qquad \Re(x) \geqq -\tfrac{1}{2};$$

$$(5.9)\qquad \left|\frac{(1+x)f(x)}{1+xf(x)} - \frac{1}{2-g_1}\right| \leqq \frac{1-g_1}{2-g_1}, \qquad \Re(x) \geqq -\tfrac{1}{2}.$$

By means of the relation (4.1) these inequalities become available for an arbitrary moment generating function.

For example, if (5.6), (5.7) are applied to the function ([2], p. 349)

$$\frac{x - \log(1+x)}{x\log(1+x)} = \frac{1}{1} + \frac{\frac{1^2}{1\cdot 2}x}{1} + \frac{\frac{1^2}{2\cdot 3}x}{1} + \frac{\frac{2^2}{3\cdot 4}x}{1} + \frac{\frac{2^2}{4\cdot 5}x}{1} + \cdots,$$

we readily obtain the simple inequality

$$(5.10)\qquad \frac{|x|}{1+|x|} \leqq |\log(1+x)| \leqq |x|\frac{|1+x|}{1+|x|}, \qquad |x| \leqq 1,$$

which is well known when x is real and $\geqq 0$.

In conclusion we shall write down the inequalities (5.6)-(5.9) for the hypergeometric function $F(\alpha, 1, \gamma; -x)$ (α, γ real and $\gamma > \alpha > 0$) which are obtained by putting

$$f(x) = \frac{1 - F(\alpha, 1, \gamma; -x)}{xF(\alpha, 1, \gamma; -x)}.$$

The g_n's are the same as the h_n's of (3.5). After simple modifications in form the inequalities become

$$(5.11)\qquad \left|F(\alpha, 1, \gamma+1; -x) - \frac{\gamma^2}{\gamma^2-\alpha^2}\right| \leqq \frac{\alpha\gamma}{\gamma^2-\alpha^2}, \qquad |x| \leqq 1;$$

$$(5.12)\qquad \left|F(\alpha, 1, \gamma+1; -x) - \frac{\gamma}{2\gamma-\alpha}\right| \leqq \frac{\gamma}{2\gamma-\alpha}, \qquad \Re(x) \geqq -\tfrac{1}{2};$$

$$(5.13)\qquad \left|\frac{1}{F(\alpha, 1, \gamma; -x)} - \frac{2\gamma-\alpha+\gamma x}{2\gamma-\alpha}\right| \leqq \frac{\gamma-\alpha}{2\gamma-\alpha}|x|, \qquad |x| \leqq 1;$$

$$(5.14)\qquad \left|\frac{1}{F(\alpha, 1, \gamma; -x)} - \frac{\gamma+\alpha+\alpha x}{\gamma+\alpha}\right| \leqq \frac{\alpha}{\gamma+\alpha}|x|, \qquad \Re(x) \geqq -\tfrac{1}{2}.$$

Bibliography

1. H. L. Garabedian and H. S. Wall, *Hausdorff methods of summation and continued fractions*, Transactions of the American Mathematical Society, vol. 48(1940), pp. 185-207.
2. O. Perron, *Die Lehre von den Kettenbrüchen*, 1st ed., 1913.
3. W. T. Scott and H. S. Wall, *A convergence theorem for continued fractions*, Transactions of the American Mathematical Society, vol. 47(1940), pp. 155-172.
4. H. S. Wall, *Continued fractions and totally monotone sequences*, Transactions of the American Mathematical Society, vol. 48(1940), pp. 165-184.

Northwestern University.

REGULAR CURVE-FAMILIES FILLING THE PLANE, I

By Wilfred Kaplan

Introduction. In studying the solutions of differential equations $dx/dt = f(x, y)$, $dy/dt = g(x, y)$ one is led to the concept of a *family of curves* in the plane.[1] Also the level-curves of a function $h(x, y)$ form a family of curves in the plane. In both cases, under proper restriction on the functions involved, the following condition, which we term *regularity*, holds: *the family is locally homeomorphic with parallel lines.*

It is the object of the present paper to consider the properties of a curve-family in the plane, given only that it satisfies the regularity condition in an open subset R of the plane. The principal theorem of the paper (see §4.2, Theorem 42 below) is that, *when R is the whole plane, the given family is a level-curve family of some function $h(x, y)$.* This generalizes a theorem of Kamke.[2]

In order to obtain this result the structural features of such a curve-family will be analyzed in some detail. In a paper to follow the structural question will be more completely considered and a topological classification of all families regular in the entire plane will be obtained.

In the first part of the paper the results of Bendixson on families defined by differential equations will be generalized to the more general (topologically defined) families. The purpose of the generalization is partly to show that Bendixson's theorems follow almost completely from the regularity alone, differentiability being inessential, and partly to express those theorems in the topological language which their nature demands. This is done particularly in the case of the discussion of the neighborhood of a closed curve (see §1.8 below). Other theorems considered are that a closed curve must enclose a singularity (see Theorem 13) and that an open curve which in one direction remains bounded and has no singularity as limit is asymptotic to a closed curve (see Theorem 11 below).

In the second part the structure of the families filling the plane will be analyzed and formulated in an algebraic manner. For such families each curve is open

Received by the Editors of the Annals of Mathematics, February 21, 1940, accepted by them and later transferred to this Journal. The material in this paper and the one to follow is taken from the author's Ph.D. thesis at Harvard University. The author expresses his gratitude to Professor Hassler Whitney for his advice in the preparation of the paper.

[1] In the bibliography at the end of this article detailed references are given to the classical works of Poincaré and Bendixson on the subject, also to works of Brouwer, v. Kerékjártó, Kneser, Denjoy, Birkhoff, Whitney, and George. Numbers in brackets refer to this bibliography.

[2] See E. Kamke, *Zur Theorie der Differentialgleichungen*, Mathematische Annalen, vol. 99(1928), p. 613.

and tends to infinity in both directions. The structure of each family can then be described in terms of the *relative positions in each triple of curves of the family.* There are essentially two types of arrangements possible for each triple. It is thus possible to regard the family as an algebraic system with two kinds of triadic order relations. Such a system we term a *chordal system.* (See the discussion in §2.2 below.)

In the third part it will be shown that each family filling the plane can be given a "normal subdivision". This subdivision is a decomposition of the family into non-overlapping subfamilies, each having the structure of the parallel lines of a half-plane (see Figure 3, §3.6). The normal subdivision makes possible the proof, in the fourth part, of the existence of a level-curve function.

1. Some fundamental theorems on regular curve-families in the plane

1.1. Families of curves. Throughout the paper our frame of reference will be the xy-plane.

DEFINITION. An *open curve* shall mean a homeomorphic image of the open interval $0 < x < 1$. A *closed curve* shall mean a homeomorphic image of the circle $x^2 + y^2 = 1$. An *arc* shall mean a homeomorphic image of the closed interval $0 \leqq x \leqq 1$. A *half-open curve* shall mean a homeomorphic image of the half-open interval $0 \leqq x < 1$. The word *curve* shall mean any one of the preceding four.

DEFINITION. A family F of curves $\{C\}$ will be said to *fill* a region R if each curve C is in R and each point of R is on one and only one curve C. If further R_0 is an open subregion of R, then F_{R_0} will denote the family of curves filling R_0 formed by the intersections of the curves C with R_0.

1.2. Regularity. (See [6], p. 241; [7], p. 136; [9], p. 249; [5], p. 462.)

DEFINITION. Let the family F of curves $\{C\}$ fill the open region R and let P be a point of R. F will be termed *regular at P* if there is an open set $U(P)$ containing P which can be mapped homeomorphically on the open rectangle $|x| < a$, $|y| < b$ in such a way that the family $F_{U(P)}$ becomes the set of lines $y =$ constant in the rectangle. F will be termed *regular in R* if it is regular at each point P of R.

THEOREM 1. *If a family F of curves C fills an open region R and is regular in R, then each curve C of F is either open or closed.*

For F could not be regular at the end-point of a half-open curve or of an arc.

THEOREM 2. *The differential equations*

$$\frac{dx}{dt} = f(x, y), \qquad \frac{dy}{dt} = g(x, y),$$

where f and g are continuous in an open region R, satisfy local Lipschitz conditions in R, and do not vanish together, determine a regular curve family which fills R.

Proof. It is known that the solutions of such differential equations form a set of trajectories $x = x(t)$, $y = y(t)$, one and only one of which passes through

each point of R. Each trajectory is a continuous one-to-one image either of an open interval or of a circle. But for the first of these two cases Bendixson ([1], p. 12) has established that the image is actually a homeomorphic one: that is, if on the curve t ranges from t_0 to t_1, then no point P of the curve is a limit point of points P_{t_n} of the curve for $\lim_{n\to\infty} t_n = t_0$ or $\lim_{n\to\infty} t_n = t_1$. In the case of a closed curve the correspondence with a circle is necessarily a homeomorphism. Hence the trajectories form a family of curves filling R.

The regularity follows from the existence of orthogonal trajectories defined by

$$\frac{dx}{dt} = -g(x, y), \qquad \frac{dy}{dt} = f(x, y).$$

The net formed by the two families of curves determines local coördinates for the mapping onto a rectangle.

1.3. **An equivalent definition of regularity.**[3] Whitney defines regularity ([9], pp. 248–249) as follows:

If H is a homeomorphism of an arc PQ on an arc $P'Q'$ such that $H(P) = P'$, $H(Q) = Q'$, let $d(H)$ denote the least upper bound of $\rho(S, H(S))$ for S on PQ and also

$$\sigma(PQ, P'Q') = \sigma(P'Q', PQ) = \sigma(QP, Q'P')$$

denote the greatest lower bound of all $d(H)$ for all possible H.

Second Definition of Regularity. A curve family F filling the open region R is *regular in* R if corresponding to each arc PQ lying on a curve C of F and each $\epsilon > 0$ there is a $\delta > 0$ such that, if P' lies on a curve C' of F and $\rho(P, P') < \delta$, there is an arc $P'Q'$ on C' such that $\sigma(P'Q', PQ) < \epsilon$. F is said to be *regular at the point* P of R if there is an open set $U(P)$ in R and containing P such that the family $F_{U(P)}$ is regular in $U(P)$.

It follows from a theorem in Whitney's paper ([9], p. 249, Theorem 7A) that F is regular in R if and only if it is regular at every point P in R.

Theorem 3. *Whitney's regularity is topologically invariant.*

To prove the theorem it must be established that if the family F_1 fills and is regular in the open region R_1 and R_1 is mapped homeomorphically on R_2, then the image family F_2 is regular in R_2.

Let P_2Q_2 be an arc lying on a curve of F. Let P_1Q_1 be its inverse image in F_1. Then, given $\epsilon > 0$, we first choose a bounded open set W_1 such that $P_1Q_1 \subset W_1 \subset \overline{W}_1 \subset R_1$, and then a positive δ_0 so small that, for S_1 and S_1' in W_1, $\rho(S_1, S_1') < \delta_0$ implies $\rho(S_2, S_2') < \epsilon$. (This is possible since $\overline{W}_1$ is compact.) Take δ_0 further so small that $V_{\delta_0}(P_1Q_1) \subset W_1$. By the regularity of F_1 we can

[3] We shall use the following notations: $\rho(x, y)$ for Euclidean distance, $V_\epsilon(A)$ for the ϵ-neighborhood of set A, $\subset$ for set inclusion, $A \cup B$ and $\sum A_i$ for set union, $A \cdot B$ and $\prod A_i$ for set intersection, $A - B$ for the difference between sets A and B, when $B \subset A$, $\bar{A}$ for topological closure.

find a $\delta_1 > 0$ such that for every P_1' on C_1' such that $\rho(P_1, P_1') < \delta_1$, there is an arc $P_1'Q_1'$ on C_1' and in R_1 such that $\sigma(P_1Q_1, P_1'Q_1') < \delta_0$.

Now let H_1 be a homeomorphism of P_1Q_1 on $P_1'Q_1'$ such that $d(H_1) < \delta_0$. Then, if H_2 is the corresponding homeomorphism of P_2Q_2 on the image $P_2'Q_2'$ of $P_1'Q_1'$, we have for each pair S_2, $H_2(S_2)$ the inequality $\rho(S_2, H_2(S_2)) < \epsilon$ and hence $\sigma(P_2Q_2, P_2'Q_2') < \epsilon$. Now choose δ so small that $\rho(P_2, P_2') < \delta$ implies $\rho(P_1, P_1') < \delta_1$. This δ has the property which implies regularity of F_2.

THEOREM 4. *The two definitions of regularity are equivalent.*

Proof. If the first condition holds, then F is locally homeomorphic with the parallel lines. But the second condition holds for the parallel lines, hence, by the preceding theorem, for the given family.

The converse is established by Whitney ([9], p. 260). Hence the two definitions are equivalent.

1.4. **Cross-sections and r-neighborhoods.** Let F be a curve family filling an open region R.

DEFINITION. The arc pq will be termed a *cross-section* relative to F if there is an open set R_0 in R containing pq and such that each curve of F_{R_0} meets pq at most once. If S is on pq, $S \neq p$, $S \neq q$, pq will be termed a *cross-section through* S.

DEFINITION. An *r-neighborhood* of a point P of R shall mean an open set $U(P)$ in R, containing P, and such that the closure $\overline{U(P)}$ can be mapped homeomorphically on the closed rectangle $|x| \leqq a$, $|y| \leqq b$ in such a way that the curves of $F_{U(P)}$ become the lines $y =$ constant, while the inverse images of the two lines $|x| = a$ are cross-sections relative to F. The homeomorphism is then said to be an *r-homeomorphism* on the *r-rectangle* $|x| \leqq a$, $|y| \leqq b$.

THEOREM 5. *If F is regular at the point P of R, then* (a) *there is an arbitrarily small r-neighborhood of P, and* (b) *there is a cross-section qs relative to F through P.*

Proof. By the regularity we can choose an open set $U_1(P)$ containing P which we can map homeomorphically on the rectangle $|x| < a$, $|y| < b$ in such a way that $F_{U_1(P)}$ is transformed onto the family of lines $y =$ constant in the rectangle. Let the image of P be (x_0, y_0). Take $\delta > 0$ and less than min $(x_0 + a, a - x_0, y_0 + b, b - y_0)$. The inverse image of $|x - x_0| \leqq \delta$, $|y - y_0| \leqq \delta$ is an r-neighborhood $U(P)$, and that of $x = x_0$, $|y - y_0| \leqq \delta$ is a cross-section qs through P. $U(P)$ can be made arbitrarily small by suitable choice of δ.

1.5. **An equivalent definition of cross-section.** Whitney gives a definition ([9], pp. 256, 260) which, in the case under discussion, is as follows:

DEFINITION. Let F be regular at P. Then a *cross-section through P* is an arc rs through P and such that

(a) each point of rs lies within an arc $Q_0'Q_1'$ lying on a curve of F such that for some $\lambda' > 0$ each arc of $N_{\lambda'}(Q_0'Q_1')$ contains at most one point of rs;

(b) P lies within an arbitrarily small arc Q_0Q_1 on a curve of F such that for some $\lambda > 0$ each arc of $N_\lambda(Q_0Q_1)$ contains exactly one point of rs.

Here $N_{\lambda'}(Q_0'Q_1')$ denotes the set of all arcs xy lying on curves of F and such that $\sigma(xy, Q_0'Q_1') < \lambda'$. It includes xy.

Using this definition, Whitney ([9], p. 260) establishes the following

LEMMA. *If rs is a cross-section through P, then there is an r-neighborhood $U(P)$ such that in a corresponding r-rectangle $|x| \leqq 1$, $|y| \leqq 1$ rs has image $x = 0$, $|y| \leqq 1$.*

By means of this we can establish

THEOREM 6. *If the curve family F fills the open region R and is regular in R, then the two definitions of cross-section are equivalent.*

Proof. Suppose rs is a cross-section through P by the second definition. Let α and β be the cross-sectional arcs of the boundary of $U(P)$ of the above lemma. Take r-neighborhoods $U(r)$ and $U(s)$ so small as not to meet α or β. The set $U(P) \cup U(r) \cup U(s)$ is an open set R_0 containing rs and such that rs meets each curve of F_{R_0} at most once. Hence rs is a cross-section by the first definition.

Conversely, suppose rs is a cross-section through P by the first definition and choose the corresponding set R_0 including rs. To each P' on rs choose an arc $Q_0'Q_1'$ in R_0. It follows from Whitney's Theorem 6B that we can further choose an $N_{\lambda'}(Q_0'Q_1')$ lying in R_0. This $N_{\lambda'}(Q_0'Q_1')$ then satisfies condition (a) of the definition. To satisfy condition (b) we choose any arc Q_0Q_1 in R_0 and through P, then an r-neighborhood $U(P)$ in R_0, not containing Q_0 or Q_1, and so small that every curve of $F_{U(P)}$ meets rs. Let $\delta_1 = \min(\rho(Q_0, \overline{U(P)}), \rho(Q_1, \overline{U(P)}))$ and choose a $V_{\delta_2}(P)$ in $U(P)$. Choose $\lambda < \delta_1$, $\lambda < \delta_2$, and so small that $N_\lambda(Q_0Q_1) \subset R_0$. Then each arc of $N_\lambda(Q_0Q_1)$ is in R_0 and meets rs just once.

1.6. **A further consequence of regularity.** In the following sections (§§1.6–1.11) F will denote a curve family which fills an open region R and is regular in R.

DEFINITION. Let C be an open curve given by $x = x(t)$, $y = y(t)$ in $0 < t < 1$. Let C be directed positively with increasing t. Let P_t denote the point $x = x(t)$, $y = y(t)$. Then a *positive limit point* of C shall mean a point Q which is the limit of a sequence of points P_{t_n} $(n = 1, 2, \cdots)$ on C such that $\lim_{n\to\infty} t_n = 1$; a *negative limit point* of C shall mean a point Q which is the limit of a sequence P_{t_n} with $\lim_{n\to\infty} t_n = 0$. In either case, Q is a *limit point* of C.[4]

THEOREM 7. *Let Q be in R and a positive (negative) limit point of the directed open curve C of F. Then every point of the curve D of F through Q is a positive (negative) limit point of C.*

[4] These limit points are the "α and ω limit points" used by Birkhoff ([2], p. 197).

D cannot coincide with C since C is a homeomorphic image of an open interval. Let Q_1 be any other point of D, QQ_1 an arc of D joining Q and Q_1. Let P_{t_0} be any point of C and let $\rho_0 = \rho(P_{t_0}, QQ_1)$. Given an $\epsilon > 0$, let $\epsilon_1 = \min(\epsilon, \rho_0)$ and $\delta = \delta(QQ_1, \epsilon_1)$ of the second condition of regularity. By assumption there is a P_{t_1} in $V_\delta(Q)$ and with $t_1 > t_0$. Hence there is an arc $P_{t_1}P_{t_2}$ of C with $\sigma(P_{t_1}P_{t_2}, QQ_1) < \epsilon_1$. Therefore $t_2 > t_0$ and $P_{t_2} \subset V_\epsilon(Q_1)$. This holds for every t_0. Hence Q_1 is a positive limit point of C. The case of a negative limit point is treated in the same way.

1.7. The neighborhood of an arc.

THEOREM 8. *Let PQ be an arc lying on a curve C of F. Then there is an r-neighborhood U containing PQ. Further U can be taken to lie within $V_\epsilon(PQ)$ for any given $\epsilon > 0$.*

Proof. Choose points P_1 and Q_1 on C so that the arc P_1Q_1 includes PQ in its interior and that P separates P_1 and Q on P_1Q_1. For each point S on PQ choose an r-neighborhood containing it and so small as to meet P_1Q_1 only in an open curve $H(S)$. By the Heine-Borel theorem, a finite number of the curves, say $H_1(S_1), H_2(S_2), \cdots, H_n(S_n)$, cover PQ. If $n = 1$, the theorem is immediate.

Suppose $n = 2$, let $U_1(P)$ and $U_2(Q)$ be the corresponding neighborhoods, $\bar{U}_1'$ and $\bar{U}_2'$ the corresponding closed r-rectangles: $\bar{U}_1'$: $|x_1'| \leqq 1$, $|y_1'| \leqq 1$; $\bar{U}_2'$: $|x_2'| \leqq 1$, $|y_2'| \leqq 1$. Let the image of $H_1 \cdot PQ$ in U_1' be $y_1' = 0$, $a \leqq x_1' < 1$, and that of $H_2 \cdot PQ$ in U_2' be $y_2' = 0$, $-1 < x_2' \leqq b$. The line $x_1' = 1$ of $\bar{U}_1'$ corresponds to a curve $x_2' = \varphi(y_2')$ in U_2' defined and continuous at least in an interval $|y_2'| \leqq \delta$, for δ sufficiently small and > 0. U_1 and U_2 can be taken to be so small that the image of $\bar{U}_1 \cdot \bar{U}_2$ in $\bar{U}_2'$ consists precisely of the set $-1 \leqq x_2' \leqq \varphi(y_2')$, $|y_2'| \leqq \delta$. The inverse image of the set $\varphi(y_2') \leqq x_2' \leqq 1$, $|y_2'| \leqq \delta$ plus the set $\bar{U}_1$ then form the closure of an r-neighborhood containing PQ.

Finally consider the general case, for any n. Suppose the proposition established for the case $n - 1$. Then, if $H_n(S_n) = H_n(Q)$, the arcs $H_1(S_1), \cdots, H_{n-1}(S_{n-1})$ cover an arc $PQ - (H_n \cdot PQ)$ and thus, by the theorem for $n - 1$, there exists a single r-neighborhood $U_1^*(P)$ covering this curve. U_1^* can then be taken smaller, if necessary, to ensure that it meets P_1Q_1 in one curve. Then $U_1^*(P)$ and $U_n(Q)$ together can be replaced by a single r-neighborhood, by the theorem for $n = 2$. Hence in all cases the theorem follows.[5]

U can always be taken smaller, if necessary, to ensure that $U \subset V_\epsilon(PQ)$.

1.8. The neighborhood of a closed curve.

Let r, θ be polar coördinates in the plane, $f(\rho)$ a continuous monotone strictly increasing function of the real variable ρ in $\rho' \leqq \rho \leqq \rho''$ with $0 < \rho' < 1 < \rho''$ and $0 < f(\rho') < 1 < f(\rho'')$. For each ρ in $\rho' < \rho < \rho''$ let α_ρ be the arc or closed curve

$$r = \rho + \frac{\theta}{2\pi}(f(\rho) - \rho) \qquad (0 \leqq \theta \leqq 2\pi).$$

[5] A similar proof is given by George [5], pp. 464–465.

Let $\rho_0 = \rho$, $\rho_n = f(\rho_{n-1})$, $\rho_{-n} = f^{-1}(\rho_{-n+1})$ $(n = 1, 2, \cdots)$, the definitions holding only as long as f or its inverse f^{-1} is defined at the value indicated. The curves α_{ρ_n} $(n = 0, \pm 1, \pm 2, \cdots)$ then together form a single curve $C_\rho = C_{\rho_n}$. In case C_ρ is half-open or an arc we agree to remove its end-points, so that in general C_ρ is an open or closed curve. The curves C_ρ for $\rho' < \rho < \rho''$ form a curve family filling the region

$$\begin{cases} \rho' + \dfrac{\theta}{2\pi}(f(\rho') - \rho') < r < \rho'' + \dfrac{\theta}{2\pi}(f(\rho'') - \rho''), \\ 0 \leqq \theta \leqq 2\pi \end{cases}$$

from which the boundary points on $\theta = 0$ have been removed. Such a family is termed a *spiral curve family* and the region which it fills a *spiral region*. The region is open.

The curves of such a family are either circles, when $f(\rho) = \rho$, or spirals asymptotic to circles.

DEFINITION. Let F be the curve family of §§1.6 and 1.7 and let C be a closed curve of F. Then an *s-neighborhood* of C will mean an open subset G of R containing C which can be mapped homeomorphically onto a spiral region G' in such a way that the curves of F_G are transformed onto the curves of a spiral curve family filling G'.

THEOREM 9. *There is an s-neighborhood of every closed curve C of F.*

Proof. Let P be a point of C and choose an r-neighborhood $U(P)$ so small that it meets C only in a single curve QS through P. The complementary arc $(QS)'$ on C is then, by Theorem 8, contained in an r-neighborhood U_1, which we can assume chosen not to include P and so small that it meets C only in one curve Q_1S_1, whereby Q, Q_1, P, S_1, S lie in that order on C.

Now construct corresponding closed r-rectangles $\bar{U}'$ and $\bar{U}_1'$. Let $\bar{U}'$ be given by $|x'| \leqq 1$, $|y'| \leqq 1$ and $\bar{U}_1'$ by $|x_1'| \leqq 1$, $|y_1'| \leqq 1$. If U_1 is chosen sufficiently small, the cross-sectional sides α_1 and β_1 of $\bar{U}_1$ have images in U' consisting respectively of curves $x' = \varphi(y')$ in $a \leqq y' \leqq b$, and $x = \psi(y')$ in $c \leqq y' \leqq d$, and images in $\bar{U}_1'$ consisting respectively of $x_1' = +1$, $x_1' = -1$. If (x_0', y_0') is the image of P in U', then we have $a < y_0' < b$, $c < y_0' < d$. The inverse images of the sets $-1 \leqq x' \leqq \varphi(y')$, $a \leqq y' \leqq b$ and $\psi(y') \leqq x' \leqq 1$, $c \leqq y' \leqq d$ then lie in $\bar{U}_1$. Let $\delta = \min(b - y_0', y_0' - a, d - y_0', y_0' - c)$. The inverse image of $\varphi(y') \leqq x' \leqq \psi(y')$, $|y' - y_0'| < \delta$ is then the closure $\bar{U}_2$ of an r-neighborhood U_2. $\bar{U}_2$ overlaps $\bar{U}_1$ only along α_1 and β_1, in fact has cross-sectional sides $\alpha_2 \subset \alpha_1$ and $\beta_2 \subset \beta_1$.

We now map $\bar{U}_2$ r-homeomorphically on the r-rectangle $1 \leqq x_2' \leqq 2$, $\frac{1}{2} \leqq y_2' \leqq \frac{3}{2}$ so that α_2 becomes $x_2' = 1$, β_2 becomes $x_2' = 2$, P has image on $y_2' = 1$. We also map $\bar{U}_1$ r-homeomorphically on the r-rectangle $0 \leqq x_2' \leqq 1$, $a_2 \leqq y_2' \leqq b_2$, with $0 < a_2 < \frac{1}{2}$, $b_2 > \frac{3}{2}$ in such a way that along α_2 these two homeomorphisms coincide. The inverse image of the set $0 \leqq x_2' \leqq 1$, $\frac{1}{2} \leqq y_2' \leqq \frac{3}{2}$ is then the closure of an r-neighborhood U_3. Now let $\rho = y_2'$, let $f(\rho)$ be the y_2'-coordinate of the point on $x_2' = 0$ with the same inverse on β_2 as the point $(2, \rho)$.

Then $f(1) = 1$ and $f(\rho)$ is defined for $\frac{1}{2} \leqq \rho \leqq \frac{3}{2}$. We now identify points on $x_2' = 0$ and $x_2' = 2$ corresponding to the same point on β_2. This can be achieved by the transformation

$$\theta = \frac{x_2'}{\pi}, \qquad r = \frac{x_2'}{2}\,(f(y_2') - y_2') + y_2'.$$

This maps the set $\bar{G} = \bar{U}_2 \cup \bar{U}_3$ homeomorphically on the closure of a spiral region as above, with $\rho' = \frac{1}{2}$, $\rho'' = \frac{3}{2}$. Each line $y_2' = \rho$ becomes an arc α_ρ and, since $f(1) = 1$, C becomes the circle $r = 1$. $\bar{G}$ is thus the closure of an s-neighborhood G of C and the theorem is established. (Cf. Bendixson [1], p. 15.)

Definition. Let the curve C of F meet the cross-section pq at points t and u interior to pq; let t and u determine arcs $(tu)_1$ on pq and $(tu)_2$ of C meeting only at t and u and hence forming a closed curve K. If further K contains neither p nor q in its interior, then K will be termed a *bay*.

Corollary to Theorem 9. *Under the assumptions of Theorem 9, if D is a curve of F which is asymptotic to C in both directions on D and D is interior to C, then an arc of D forms part of a bay within C.*

Proof. D cannot lie wholly in an s-neighborhood G of C. Take a point P on D and not in G. Choose a cross-section rs in G and such that r is interior to C, s is exterior to C, and that r is not on D. Let t and u be the first points at which D meets rs in the two directions on D from P. Let K be the closed curve formed by the cross-section tu plus the arc tPu of C. In the homeomorphic spiral curve family, the image of $D \cdot G$ must consist of two spirals approaching the circular image of C, both in the clockwise direction, or both in the counterclockwise direction. This implies that r and s can be joined by an arc in G and not meeting K. But K is within C, hence s is exterior to K, and, accordingly, r is exterior to K. Thus K is a bay within C.

Theorem 10. *Let C be a curve of F asymptotic to the closed curve D of F in one direction. Then to every point P of C corresponds a neighborhood $V_\delta(P)$ so small that every curve of F crossing $V_\delta(P)$ is asymptotic to D in one direction.*[6]

Proof. The theorem certainly holds if R is a spiral region and hence in general if P is in an s-neighborhood of D. If P is not in an s-neighborhood G of D, choose a point P_1 of C in G and a neighborhood $V_{\delta_1}(P_1)$ with the property desired for P_1. By the second condition of regularity a neighborhood $V_\delta(P)$ can be found such that all curves of F crossing $V_\delta(P)$ cross $V_{\delta_1}(P_1)$, hence are asymptotic to D.

1.9. Curves bounded in one direction.

Theorem 11. *Let C be a directed open curve of F which is bounded in the positive direction but has no boundary point of R as a positive limit point. Then C is asymptotic to a closed curve in the positive direction on C.*[7]

[6] Cf. Bendixson [1], p. 15.
[7] Cf. Bendixson [1], p. 11.

Proof. Since C is bounded in the given direction, it must have limit points. Since C is topological image of an open interval, no limit point of C can lie on C. By Theorem 7, the set E of all positive limit points of C is a union of curves of F, and since C is bounded in the given direction, E must also be bounded. Any limit point of E belongs to E, hence E is a closed bounded set, therefore bicompact.[8] Suppose E contains no closed curve of F.

Let C_1 be a curve of E. Let E_1 be the set of limit points of C_1. E_1 is non-void, is a closed set, a subset of E and a union of curves of F. Let C_2 be a curve of E_1 and E_2 the set of limit points of C_2. Then E_2 is a non-void closed subset of E_1. Proceeding in this way, we obtain a sequence $E \supset E_1 \supset E_2 \supset E_3 \supset \cdots$ of closed sets, each of which is a union of open curves of F and, moreover, E_{n+1} is a proper subset of E_n.

Suppose now that we have defined the sets E for all ordinal numbers $\alpha < \beta$ in such a way that $\alpha < \alpha'$ implies $E_\alpha \supset E_{\alpha'}$, each E_α is a non-void closed set and is the union of open curves of F. Then set $E_\beta^* = \prod_{\alpha<\beta} E_\alpha$. Since the sets E_α form a well-ordered monotone decreasing sequence of non-void closed sets, their intersection E_β^* is non-void,[9] is a closed set, and is also a union of open curves of F. Now take a curve C_β^* in E_β^* and let E_β be the set of limit points of C_β^*. E_β is then a proper non-void closed subset of E_β^*, hence of all E_α for $\alpha < \beta$. Thus we have a transfinite sequence E_α defined which is actually decreasing. But it follows from the Baire-Hausdorff theorem[10] that only a countable number of the sets E_α can be distinct, and this is a contradiction. Hence the theorem follows.

1.10. The interior of closed curves.

THEOREM 12. *Interior to a bay there is a boundary point of R.*

Proof. Let the bay be a closed curve K formed by an arc tPu of a curve C of F and a cross-section tu within a cross-section rs. By the lemma of §1.5 there is an r-neighborhood U on whose boundaries r and s lie, while the rest of rs is interior to U. Let α and β be the cross-sectional sides of $\bar{U}$. Since r and s are exterior to K, we can assume that K meets α and not β. We can assume $P \not\subset \bar{U}$, $\rho(P, \bar{U}) = \eta > 0$.

Let $\epsilon_1 = \rho(tPu, \beta)$, $\epsilon_2 = \rho(u, \bar{U} - U)$, $\epsilon = \min(\epsilon_1, \epsilon_2, \eta)$, and choose $\delta = \delta(\epsilon, tPu)$ of the second condition of regularity. Every curve C_1 of F crossing rs at t_1 in $V_\delta(t)$ must then contain an arc t_1u_0 with $\sigma(t_1u_0, tPu) < \epsilon$. t_1u_0 cannot cross β and u_0 is in U. Hence C_1 enters the interior of K at t_1 and then meets rs again at a point u_1. Suppose that every curve of F crossing tu at a point t_1 were to enter the interior of K and leave at u_1 on tu. Then, as t_1 moved continuously towards u along tu, u_1 would move continuously towards t. At some intermediate point t_1 and u_1 coincide, and this is impossible.

[8] See Alexandroff-Hopf, *Topologie* I, Berlin, 1935, pp. 86, 88.

[9] See P. Alexandroff and P. Urysohn, *Zur Theorie der topologischen Räume*, Mathematische Annalen, vol. 91(1924), pp. 258-266.

[10] See Alexandroff-Hopf, loc. cit. (footnote 8), p. 79.

There must then be a first point t^* on tu such that the curve C^* through t^* enters the interior of K in one direction but fails to leave. Suppose now that there is no boundary point inside K. Then by Theorem 11, C^* is asymptotic to a closed curve D within K, and by Theorem 10 the same holds for curves crossing tu near t^*. This contradicts our assumption that t^* was the first such point. Hence there is a boundary point inside K.

Remark. Bays cannot occur in families defined by differential equations. But the regular curve family consisting of a set of confocal ellipses plus the limiting line segment minus its end-points does contain bays.

THEOREM 13. *Let C be a closed curve of F. Then interior to C there is a boundary point of R.*[11]

Proof. Suppose there is no boundary point interior to C. Let D be a curve of F interior to C. Then D is bounded completely, hence is either a closed curve or, by Theorem 11, is asymptotic to a closed curve. If D were asymptotic to just one closed curve, then by the corollary to Theorem 9 and by Theorem 12, there is a boundary point inside C, contrary to assumption. Hence there must be closed curves (and, by the same reasoning, infinitely many) inside C.

Now for every point P inside C set $A(P)$ = the area enclosed by the curve D through P if D is closed, and set $A(P) = B$ = the area enclosed by C if D is not closed. Set g = g.l.b. $A(P)$ for P inside C. Then from the preceding paragraph we have $g < B$.

Let $\{P_n\}$ be a sequence of points inside C such that $\lim_{n\to\infty} A(P_n) = g$. Let P be a limit point of $\{P_n\}$.

Suppose P lies on a closed curve D of F. Then there must be another closed curve of F inside D, whence $A(P) > g$. Further take an s-neighborhood G of D and a neighborhood $V_\delta(P)$ therein. The curves of G are either spirals or closed curves. Moreover, the closed curves are concentric (in the sense that their interiors can be simply ordered with respect to inclusion) since they are the images of the concentric circles of a spiral curve family. If a closed curve D_1 crosses $V_\delta(P)$ at a point Q_1 and lies within D, then $A(Q_1) = g_1 > g$. For any other closed curve D_2 lying between D_1 and D and crossing $V_\delta(P)$ at Q_2, $A(P) > A(Q_2) > g_1$. For any closed curve D meeting $V_\delta(P)$ at Q_3 and including D in its interior $A(Q_3) > A(P) > g_1$. For any spiral crossing $V_\delta(P)$ at Q_4, $A(Q_4) = B > g_1$. Thus in $V_\delta(P)$, $A(P) > g_1 > g$. This contradicts our assumption that P is a limit point of $\{P_n\}$.

On the other hand, if P lies on a spiral, then by Theorem 10 all curves crossing a neighborhood $V_\delta(P)$ are spirals. Hence in that neighborhood $A(Q) = B$. This is again a contradiction. Hence the theorem follows.

Remark. This proof is based on that of Bendixson.

[11] Cf. Bendixson [1], p. 16.

2. Regular curve-families filling the plane and the corresponding chordal systems

2.1. Fundamental properties. From this point on we shall assume that the curve-family F is regular and fills the plane. It may also be pictured as a regular family filling the surface of the sphere with the exception of one singular point or as a family filling a simply-connected open region of the plane.

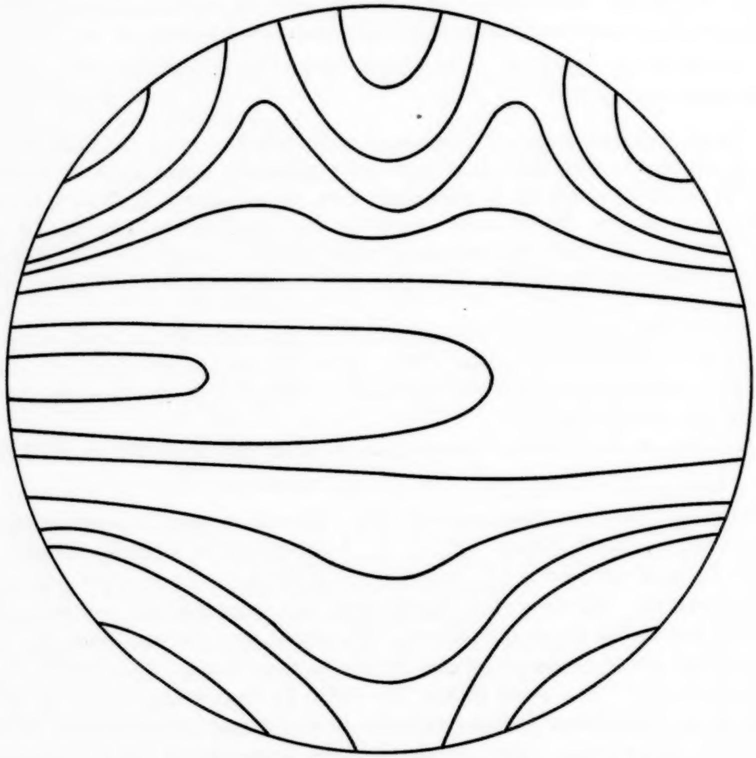

FIG. 1. EXAMPLE OF A REGULAR CURVE-FAMILY FILLING THE PLANE
The plane is represented homeomorphically as the interior of the circle.

THEOREM 14. *Each curve of F is open.*

This follows from Theorems 1 and 13.

THEOREM 15. *An arc rs is a cross-section relative to F if and only if it meets each curve of F at most once.*

Proof. If rs meets each curve of F at most once, then rs is a cross-section, as follows from the definition (see §1.4) if we take R_0 as the plane. Conversely,

suppose rs is a cross-section and that a curve C_1 crosses rs at successive points r_1 and s_1. Then the arc $(r_1s_1)_1$ of rs plus the arc $(r_1s_1)_2$ of C_1 forms a closed curve K. Let r_2 be a point on $(r_1s_1)_1$ not an end-point and not a point of C_1. The curve C_2 through r_2 then enters the interior of K at r_2, but by Theorems 11 and 13 must leave again at s_2 on $(r_1s_1)_1$, thus forming a bay. This is impossible by Theorem 12. Hence each curve of F meets rs at most once.

THEOREM 16. *Each curve C of F tends to infinity in both directions, i.e., has infinity as sole limit point.*

Proof. If C had a finite limit point Q, it would meet a cross-section through Q infinitely often, and this contradicts Theorem 15.

In Figure 1 is illustrated a regular family filling the interior of a circle (homeomorphic image of the plane).

2.2. Algebraic formulation. Chordal systems. For the following discussion we use solely the following feature of F: it is a collection of open curves, no two of which intersect, each one of which tends to infinity in both directions.

It follows from Jordan's theorem (as applied to the sphere) that each curve C of F divides the plane into two distinct regions, of which C is the common boundary. If we now consider three distinct curves C_1, C_2, C_3 of F, we see intuitively that there are actually five different ways in which they can divide the plane. We exemplify these by figures (Figure 2).

With each triple of curves may thus be associated a certain interrelationship. We shall later rigorously formalize this as follows: for case (a), we write $C_1 \mid C_2 \mid C_3$; for case (b), $C_3 \mid C_1 \mid C_2$; for case (c), $C_2 \mid C_3 \mid C_1$; for case (d), $\mid C_1, C_2, C_3 \mid^+$; for case (e), $\mid C_1, C_2, C_3 \mid^-$. In the first three cases the symbol indicates which curve is in the middle position. The symbols $+$ and $-$ correspond respectively to counterclockwise and clockwise orientation.

We note further that a number of laws hold connecting the different relations. For example, if $C_1 \mid C_2 \mid C_3$, then $C_3 \mid C_2 \mid C_1$. If $\mid C_1, C_2, C_3 \mid^+$, then $\mid C_2, C_3, C_1 \mid^+$, while $\mid C_1, C_3, C_2 \mid^-$. We can obtain many such laws, and this suggests regarding the curve family as a certain sort of algebraic system in which essentially two different relations are defined and obey certain laws. Such a system, which we shall shortly define precisely, we shall call a *chordal system*. Each curve family determines such a system in a unique way. The term chordal system is used because a set of non-intersecting chords on a circle forms such a system. Moreover, as will be shown in a later paper, each curve-family can be represented in a significant way by such a set of chords.

2.3. Definition of abstract chordal systems. Let E be a non-void class of elements $a, b, c, \cdots$. E is called a chordal system, or a CS, if there is a set of two relations $a \mid b \mid c$ and $\mid a, b, c \mid^+$ defined for the different triples of distinct elements a, b, c in E, in such a way that the following axioms hold:

AXIOM 1. *For each triple a, b, c of distinct elements one and only one of the following five relations holds*: $a \mid b \mid c$, $b \mid c \mid a$, $c \mid a \mid b$, $\mid a, b, c \mid^+$, $\mid a, c, b \mid^+$.

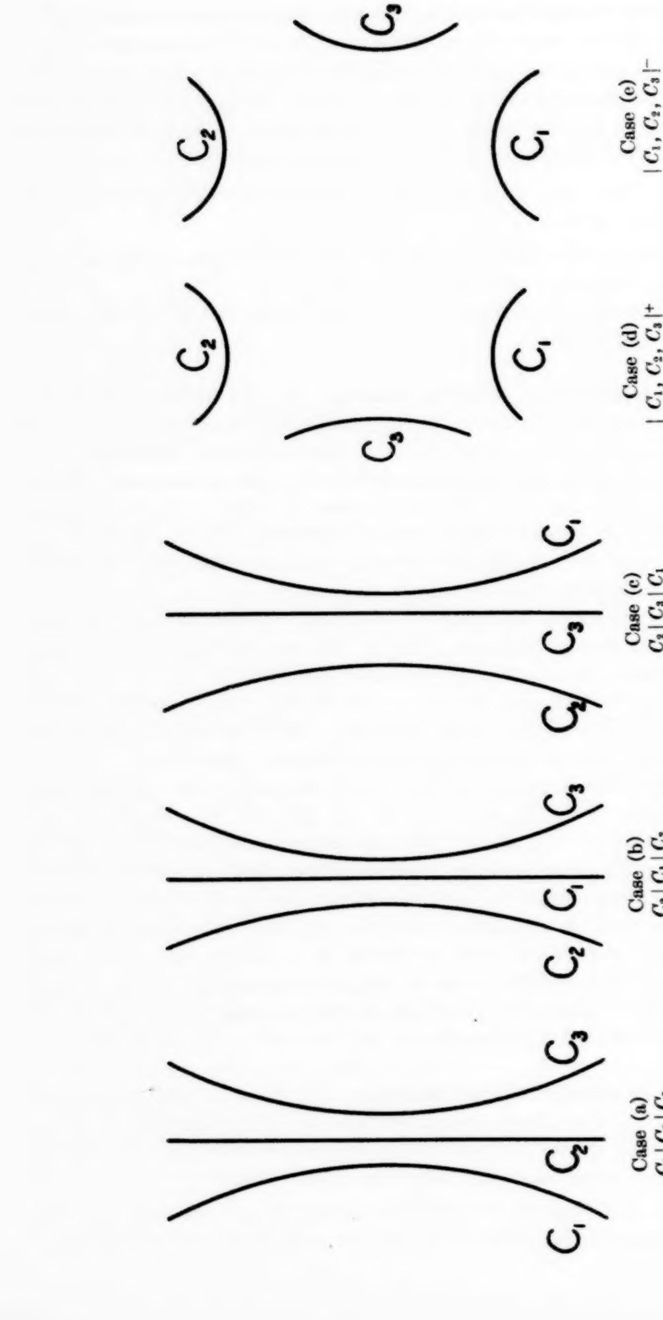

FIG. 2. POSSIBLE PLACINGS OF THREE CURVES

Notation. $| a, b, c |^-$ shall mean $| a, c, b |^+$. $| a, b, c |^{\pm}$ shall mean $| a, b, c |^+$ or $| a, b, c |^-$. $[a, b, c] \sim [a', b', c']$ shall mean that the relation holding for a, b, c holds for a', b', c' when a is replaced by a', b by b', c by c'.

AXIOM 2.1. *$a \mid b \mid c$ is equivalent to $c \mid b \mid a$.*

AXIOM 2.2. *$| a, b, c |^+$ is equivalent to $| b, c, a |^+$ (and hence to $| c, a, b |^+$).*

AXIOM 3.1. *$| a, b, c |^+$ and $| a, c, d |^+$ imply $| a, b, d |^+$ and $| b, c, d |^+$.*

AXIOM 3.2. *$| a, b, c |^{\pm}$ and $a \mid b \mid d$ imply $c \mid b \mid d$ and $| a, d, c |^{\pm}$, whereby $[a, b, c] \sim [a, d, c]$.*

AXIOM 3.3. *$a \mid b \mid c$ and $b \mid c \mid d$ imply $a \mid b \mid d$ and $a \mid c \mid d$.*

AXIOM 3.4. *Of the three relations $b \mid a \mid c$, $b \mid a \mid d$ and $c \mid a \mid d$ at most two can hold.*

THEOREM 17. *$a \mid b \mid c$ and $a \mid c \mid d$ imply $a \mid b \mid d$ and $b \mid c \mid d$.*

Proof. By virtue of Axioms 1, 2.1, and 2.2, there are twenty-five different possibilities for the second pair of triples. $d \mid a \mid b$ and $a \mid b \mid c$ imply by Axiom 3.2 $d \mid a \mid c$. Hence, by Axiom 1, (1) $d \mid a \mid b$ is false. $| a, b, d |^{\pm}$ and $a \mid b \mid c$ imply by Axiom 3.2 $| a, c, d |^{\pm}$. Hence, by Axiom 1, (2) $| a, b, d |^{\pm}$ is false. $a \mid c \mid d$ and $c \mid d \mid b$ imply by Axiom 3.3 $a \mid c \mid b$. Hence, by Axioms 1 and 2.1, (3) $c \mid d \mid b$ is false. $| b, c, d |^{\pm}$ and $a \mid b \mid c$ imply by Axiom 3.3 $| a, c, d |^{\pm}$. Hence, by Axiom 1, (4) $| b, c, d |^{\pm}$ is false. $c \mid b \mid d$ and $a \mid b \mid d$ together contradict $a \mid b \mid c$ by Axiom 3.4. Hence (5) not both $c \mid b \mid d$ and $a \mid b \mid d$ hold. $c \mid b \mid d$ and $b \mid d \mid a$ imply by Axiom 3.3 $c \mid d \mid a$. Hence, by Axiom 1, (6) not both $c \mid b \mid d$ and $b \mid d \mid a$ hold. $b \mid c \mid d$ and $a \mid b \mid c$ imply by Axiom 3.3, $a \mid b \mid d$. Hence, by Axioms 1 and 2.1, (7) not both $b \mid c \mid d$ and $a \mid d \mid b$ hold. The results (1), $\cdots$, (7) together with Axioms 2.1 and 2.2 eliminate all cases except $a \mid b \mid d$ and $b \mid c \mid d$. Hence the theorem follows.

2.4. **The relation $C_1 \mid C_2 \mid C_3$ for curves of F.** If C is a curve of F, we shall denote by $\mathfrak{D}(C)$ one of the two regions into which C divides the plane. When the choice of $\mathfrak{D}(C)$ is fixed, we shall denote by $\mathfrak{D}^*(C)$ the other region. In both cases only inner points will be included.

If C_1 and C_2 are two different curves of F, then for suitable choices of $\mathfrak{D}(C_1)$ and $\mathfrak{D}(C_2)$ respectively, $C_1 \subset \mathfrak{D}(C_2)$ and $C_2 \subset \mathfrak{D}(C_1)$. Let P be a point of $\mathfrak{D}^*(C_2)$. P must also be in $\mathfrak{D}(C_1)$. For, since $C_2 \subset \mathfrak{D}(C_1)$, if P is not in $\mathfrak{D}(C_1)$, P can be joined to any point on C_1 by an arc not meeting C_2. This implies that $C_1 \subset \mathfrak{D}^*(C_2)$, and this is impossible. Hence $\mathfrak{D}^*(C_2) \subset \mathfrak{D}(C_1)$ and similarly $\mathfrak{D}^*(C_1) \subset \mathfrak{D}(C_2)$. We summarize this situation in the following

THEOREM 18. *If C_1 and C_2 are two distinct curves of F, then $\mathfrak{D}(C_1)$ and $\mathfrak{D}(C_2)$ can be chosen in one and only one way so that each of the following conditions holds:*

(1) $C_1 \subset \mathfrak{D}(C_2)$, $C_2 \subset \mathfrak{D}(C_1)$,

(2) $\mathfrak{D}^*(C_1) \subset \mathfrak{D}(C_2)$,

(3) $\mathfrak{D}^*(C_2) \subset \mathfrak{D}(C_1)$,

(4) $\mathfrak{D}^*(C_1) \cdot \mathfrak{D}^*(C_2) = 0$.

Further, the four conditions are equivalent.

The equivalence of the four conditions is immediately verifiable. Thus suppose (4) holds. Then we must have $\mathfrak{D}(C_2) \supset \mathfrak{D}^*(C_1)$, which is (2), and similarly for the other implications.

DEFINITION 1. We write $C_1 \mid C_2 \mid C_3$ if $\mathfrak{D}(C_1)$, $\mathfrak{D}(C_2)$, and $\mathfrak{D}(C_3)$ can be so chosen that $\mathfrak{D}(C_1) \subset \mathfrak{D}(C_2) \subset \mathfrak{D}(C_3)$.

THEOREM 19. *Definition* 1 *satisfies Axioms* 2.1, 3.3, *and* 3.4.

Proof. We first prove that Definition 1 satisfies Axiom 2.1. By Theorem 18, $\mathfrak{D}(C_1) \subset \mathfrak{D}(C_2) \subset \mathfrak{D}(C_3)$ is equivalent to $\mathfrak{D}^*(C_3) \subset \mathfrak{D}^*(C_2) \subset \mathfrak{D}^*(C_1)$. Hence $C_1 \mid C_2 \mid C_3$ is equivalent to $C_3 \mid C_2 \mid C_1$.

Next we show that Definition 1 satisfies Axiom 3.3. Suppose $C_1 \mid C_2 \mid C_3$ and $C_2 \mid C_3 \mid C_4$. Then $\mathfrak{D}(C_1) \subset \mathfrak{D}(C_2) \subset \mathfrak{D}(C_3)$ for proper choices and $\mathfrak{D}'(C_2) \subset \mathfrak{D}'(C_3) \subset \mathfrak{D}'(C_4)$ for second proper choices $\mathfrak{D}'(C_i)$ of the $\mathfrak{D}(C_i)$ $(i = 2, 3, 4)$. From Theorem 18, we must have $\mathfrak{D}'(C_2) = \mathfrak{D}(C_2)$, $\mathfrak{D}'(C_3) = \mathfrak{D}(C_3)$. Hence $\mathfrak{D}(C_1) \subset \mathfrak{D}(C_2) \subset \mathfrak{D}'(C_4)$ and $\mathfrak{D}(C_1) \subset \mathfrak{D}(C_3) \subset \mathfrak{D}'(C_4)$. Accordingly, $C_1 \mid C_2 \mid C_4$ and $C_1 \mid C_3 \mid C_4$.

Finally to show that Definition 1 satisfies Axiom 3.4, suppose $C_2 \mid C_1 \mid C_3$, $C_2 \mid C_1 \mid C_4$, and $C_3 \mid C_1 \mid C_4$. Then (1) $\mathfrak{D}(C_2) \subset \mathfrak{D}(C_1) \subset \mathfrak{D}(C_3)$, (2) $\mathfrak{D}'(C_2) \subset \mathfrak{D}'(C_1) \subset \mathfrak{D}'(C_4)$, (3) $\mathfrak{D}''(C_3) \subset \mathfrak{D}''(C_1) \subset \mathfrak{D}''(C_4)$ for proper choices. By Theorem 18, (1) and (2) imply $C_2 \subset \mathfrak{D}(C_1)$, $C_3 \subset \mathfrak{D}^*(C_1)$, $C_4 \subset \mathfrak{D}^*(C_1)$, while (3) implies that C_3 and C_4 cannot both be in $\mathfrak{D}^*(C_1)$. This is a contradiction.

2.5. The relation $\mid C_1, C_2, C_3 \mid^+$.

DEFINITION 2. We write $\mid C_1, C_2, C_3 \mid^+$ if there is a closed curve $P_1P_2P_3P_1$ meeting the curves C_1, C_2, C_3 only at the points P_1 on C_1, P_2 on C_2, P_3 on C_3 and such that the orientation $P_1P_2P_3P_1$ is positive.[12]

THEOREM 20. *Axiom* 2.2 *holds.*

Proof. The orientations $P_1P_2P_3P_1$ and $P_2P_3P_1P_2$ of the closed curve are the same. Hence $\mid C_1, C_2, C_3 \mid^+$ is equivalent to $\mid C_2, C_3, C_1 \mid^+$.

DEFINITION. An *o-homeomorphism* shall mean an orientation-preserving homeomorphism.[12]

THEOREM 21. *Let C_i $(i = 1, 2, \cdots, n)$ be n distinct curves of F such that $\mathfrak{D}^*(C_i) \cdot \mathfrak{D}^*(C_j) = 0$ for $i \neq j$ for proper choice of the $\mathfrak{D}(C_i)$. Then the set $\prod_1^n \mathfrak{D}(C_i) \cup \sum_1^n C_i$ can be mapped o-homeomorphically on the set consisting of the interior of a circle plus n distinct circular arcs C_i' minus their end-points, each arc C_i' being the image of C_i.*

[12] For full details on orientation and orientation-preserving transformations see Alexandroff-Hopf, loc. cit. (footnote 8), pp. 165-166, pp. 462-464, pp. 474-476. Throughout we assume the plane oriented so that the positive direction on a closed curve is the counterclockwise one.

Proof. The set $D = \prod_1^n \mathfrak{D}(C_i)$ is an open simply-connected set with boundary $\sum_1^n C_i$, as follows by repeated application of the Jordan curve theorem. By the Osgood-Carathéodory theorem[13] on conformal mapping, D can be mapped o-homeomorphically on the interior of the circle and the same map remains one-to-one and continuous along the boundary of D. Thus each C_i becomes a circular arc C_i' minus its end-points.

THEOREM 22. *Axiom* 1 *holds.*

Proof. Suppose $\mathfrak{D}(C_1)$ and $\mathfrak{D}(C_2)$ chosen as in Theorem 18 and that no one of the relations $C_1 \mid C_2 \mid C_3$, $C_2 \mid C_3 \mid C_1$, $C_3 \mid C_1 \mid C_2$ holds. This means $C_3 \not\subset \mathfrak{D}^*(C_1)$, $C_3 \not\subset \mathfrak{D}^*(C_2)$, and hence $C_3 \subset \mathfrak{D}(C_1) \cdot \mathfrak{D}(C_2)$. Choose $\mathfrak{D}(C_3) \supset C_1$. Hence $\mathfrak{D}^*(C_1) \subset \mathfrak{D}(C_3)$. If further $\mathfrak{D}(C_3) \subset \mathfrak{D}(C_2)$, then $\mathfrak{D}^*(C_1) \subset \mathfrak{D}(C_3) \subset \mathfrak{D}(C_2)$, and hence $C_1 \mid C_3 \mid C_2$, contrary to assumption. Hence $\mathfrak{D}^*(C_3) \subset \mathfrak{D}(C_2)$ and $\mathfrak{D}^*(C_3) \cdot \mathfrak{D}^*(C_2) = 0$. Further $\mathfrak{D}^*(C_1) \cdot \mathfrak{D}^*(C_3) = 0$ and $\mathfrak{D}^*(C_1) \cdot \mathfrak{D}^*(C_2) = 0$.

We can now apply Theorem 21 to the set $\prod_1^3 \mathfrak{D}(C_i) \cup \sum_1^3 C_i$. Join the mid-points P_1', P_2', P_3' of the arcs C_i' to obtain a triangle $P_1'P_2'P_3'$. The inverse image of this triangle is a closed curve $P_1P_2P_3$ meeting C_1, C_2, C_3 only at P_1, P_2, P_3. Either $P_1P_2P_3P_1$ or $P_1P_3P_2P_1$ is positively oriented. Hence either $\mid C_1, C_2, C_3 \mid^+$ or $\mid C_1, C_3, C_2 \mid^+$. Thus at least one of the five relations must hold for the triple C_1, C_2, C_3.

No more than one relation can hold. For if both $C_1 \mid C_2 \mid C_3$ and $C_3 \mid C_1 \mid C_2$, then $\mathfrak{D}(C_1) \subset \mathfrak{D}(C_2) \subset \mathfrak{D}(C_3)$ and $\mathfrak{D}'(C_3) \subset \mathfrak{D}'(C_1) \subset \mathfrak{D}'(C_2)$, whence $\mathfrak{D}'(C_3) = \mathfrak{D}^*(C_3)$, $\mathfrak{D}'(C_1) = \mathfrak{D}^*(C_1)$, $\mathfrak{D}'(C_2) = \mathfrak{D}^*(C_2)$. This gives $\mathfrak{D}^*(C_1) \subset \mathfrak{D}^*(C_2)$, a result which is impossible. Similarly the pairs of relations $C_1 \mid C_2 \mid C_3$ and $C_1 \mid C_3 \mid C_2$, $C_3 \mid C_1 \mid C_2$ and $C_1 \mid C_3 \mid C_2$ are impossible.

If $\mid C_1, C_2, C_3 \mid^+$, then choose the curve $P_1P_2P_3P_1$ of Definition 2. Choose $\mathfrak{D}(C_2)$ to include C_1 and it will then include C_3. If now also $C_1 \mid C_2 \mid C_3$, then $\mathfrak{D}'(C_1) \subset \mathfrak{D}'(C_2) \subset \mathfrak{D}'(C_3)$ for proper choices. Hence $C_1 \subset \mathfrak{D}'(C_2)$, $C_3 \subset \mathfrak{D}'^*(C_2)$, and this is impossible. Similarly $\mid C_1, C_2, C_3 \mid^+$ and $C_1 \mid C_3 \mid C_2$ or $C_3 \mid C_1 \mid C_2$ gives a contradiction.

If both $\mid C_1, C_2, C_3 \mid^+$ and $\mid C_1, C_3, C_2 \mid^+$, then there are closed curves $P_1P_2P_3P_1$ and $\bar{P}_1\bar{P}_3\bar{P}_2\bar{P}_1$ positively oriented as in Definition 2. We can then choose $\mathfrak{D}(C_1)$, $\mathfrak{D}(C_2)$, $\mathfrak{D}(C_3)$ as in the preceding paragraph so that $C_2 \cup C_3 \subset \mathfrak{D}(C_1)$, $C_1 \cup C_2 \subset \mathfrak{D}(C_3)$, $C_1 \cup C_3 \subset \mathfrak{D}(C_2)$. Hence by Theorem 18 $\mathfrak{D}^*(C_i) \cdot \mathfrak{D}^*(C_j) = 0$ for $i \neq j$. If we then apply Theorem 21, $P_1P_2P_3P_1$ becomes a positively oriented closed curve $P_1'P_2'P_3'P_1'$ lying in the interior of the circle except for the points P_1' on C_1', P_2' on C_2', P_3' on C_3'. This implies that the arcs C_1', C_2', C_3' in that order follow the positive orientation on the circle. It is thus impossible for $\bar{P}_1\bar{P}_3\bar{P}_2\bar{P}_1$ also to be positively oriented. Hence not both $\mid C_1, C_2, C_3 \mid^+$ and $\mid C_1, C_3, C_2 \mid^+$. The theorem is now established.

[13] See C. Carathéodory, Mathematische Annalen, vol. 73(1918), pp. 306–320.

COROLLARY. $| C_1, C_2, C_3 |^{\pm}$ *is equivalent to the condition* $\mathfrak{D}^*(C_i)\cdot\mathfrak{D}^*(C_j) = 0$, $i \neq j$, *for proper choices.*

2.6. Verification of other axioms.

THEOREM 23. *The definitions satisfy Axiom* 3.1.

Proof. Suppose $| C_1, C_2, C_3 |^+$ and $| C_1, C_3, C_4 |^+$. Then $\mathfrak{D}^*(C_i)\cdot\mathfrak{D}^*(C_j) = 0$ $(i \neq j; i, j = 1, 2, 3)$, $\mathfrak{D}'^*(C_i)\cdot\mathfrak{D}'^*(C_j) = 0$ $(i \neq j; i, j = 1, 3, 4)$ by the above corollary. We conclude from Theorem 18 that $\mathfrak{D}'^*(C_i) = \mathfrak{D}^*(C_i)$ $(i = 1, 2, 3)$. We now apply Theorem 21 to the set $\prod_1^4 \mathfrak{D}(C_i) \cup \sum_1^4 C_i$. The arcs C_1', C_2', C_3' in that order must then lie in the positive direction on the circumference and the same holds for C_1', C_3', C_4'. Hence C_1', C_2', C_3', C_4' lie in the positive order. Let the mid-points of these arcs be P_i' $(i = 1, 2, 3, 4)$. The triangles $P_1'P_2'P_4'$ and $P_2'P_3'P_4'$ then have inverses which are positively oriented closed curves. Hence $| C_1, C_2, C_4 |^+$ and $| C_2, C_3, C_4 |^+$.

THEOREM 24. *The definitions satisfy Axiom* 3.2.

Proof. Suppose $| C_1, C_2, C_3 |^+$ and $C_1 | C_2 | C_4$. Then we can choose $\mathfrak{D}(C_1)$, $\mathfrak{D}(C_2)$, and $\mathfrak{D}(C_3)$ so that $\mathfrak{D}(C_i)\cdot\mathfrak{D}(C_j) = 0$ $(i \neq j; i, j = 1, 2, 3)$. For second choices $\mathfrak{D}'(C_i)$ we have $\mathfrak{D}'(C_1) \subset \mathfrak{D}'(C_2) \subset \mathfrak{D}'(C_4)$. It follows from Theorem 18 that $\mathfrak{D}'(C_1) = \mathfrak{D}(C_1)$ and $\mathfrak{D}'(C_2) = \mathfrak{D}^*(C_2)$. Also $\mathfrak{D}(C_2)\cdot\mathfrak{D}(C_3) = 0$ implies $\mathfrak{D}(C_3) \subset \mathfrak{D}^*(C_2)$. Hence $\mathfrak{D}(C_3) \subset \mathfrak{D}^*(C_2) \subset \mathfrak{D}'(C_4)$ and $C_3 | C_2 | C_4$. Further $\mathfrak{D}(C_3)\cdot\mathfrak{D}'^*(C_4) = 0$ and $\mathfrak{D}(C_1)\cdot\mathfrak{D}'^*(C_4) = 0$, whence $| C_1, C_3, C_4 |^{\pm}$. If $| C_1, C_3, C_4 |^+$, then by the preceding theorem $| C_4, C_1, C_2 |^+$, and this is impossible. Hence $| C_1, C_3, C_4 |^-$ and $| C_1, C_4, C_3 |^+$.

We conclude, on the basis of Theorems 19, 20, 22, 23, 24,

THEOREM 25. *Let F be a regular curve-family filling the plane or, more generally, a set of non-intersecting open curves tending to* ∞ *in both directions in the plane. Then under Definitions* 1 *and* 2 *F becomes a chordal system CS(F) whose elements are the curves of F.*

3. Curve families as normal chordal systems

3.1. The idea of normal subdivision. The simplest curve family is that of the parallel lines. In every regular curve family we can extract subfamilies homeomorphic with the parallel lines: e.g., the subfamily meeting a single cross-section. (See Theorem 30 below.) It is therefore natural to try to decompose the whole family into such pieces. The decomposition is made simpler if a larger piece is used, formed by the curves crossing an infinite cross-section which starts at a finite point and runs off to infinity. Such a family is homeomorphic with the parallel lines filling a half-plane. The normal subdivision of a curve family consists precisely in the decomposition of the curve family into (in general countably many) such pieces which do not overlap. The pieces fit together in a somewhat complicated fashion, most easily described by means of the chordal relations. We shall therefore begin with a discussion of the structure of a chordal system, by means of which the normal subdivision is defined. Then,

by means of several preparatory theorems, the normal subdivision of a regular curve family will be established.

3.2. **Subdivision of a chordal system.** We shall first point out some elementary theorems on how a chordal system can be subdivided. A fixed chordal system E will be assumed.

DEFINITION. An element a of E will be said to *divide* E if there exist two disjoint subsets $\delta(a)$ and $\delta^*(a)$ of E, not containing a and such that $E = \delta(a) \cup \delta^*(a) \cup a$ and further: for b in $\delta(a)$, c in $\delta^*(a)$, $b \mid a \mid c$; conversely, if $b \mid a \mid c$, then either $b \subset \delta(a)$ and $c \subset \delta^*(a)$ or else $b \subset \delta^*(a)$, $c \subset \delta(a)$.

THEOREM 26. *Every element a of E divides E in a unique way.*

Proof. Let b_0 be any other element of E. Let $\delta^*(a)$ be the set of all elements c of E such that $b_0 \mid a \mid c$, and let $\delta(a)$ be $E - (\delta^*(a) \cup a)$. $\delta(a)$ and $\delta^*(a)$ are disjoint and cannot contain a.

Suppose b is in $\delta(a)$, $b \neq b_0$, and c is in $\delta^*(a)$. Then $b_0 \mid a \mid c$ and $b_0 \mid a \mid b$ is false. We wish to show that $b \mid a \mid c$. $b_0 \mid b \mid a$ and $b_0 \mid a \mid c$ imply by Theorem 17 $b \mid a \mid c$. $b_0 \mid b \mid a$ and $b_0 \mid a \mid c$ imply by Axiom 3.3 $b \mid a \mid c$. $\mid b_0, a, b \mid^{\pm}$ and $b_0 \mid a \mid c$ imply by Axiom 3.2 $b \mid a \mid c$. Hence in all cases $b \mid a \mid c$. If $b = b_0$, the same holds by the definition of $\delta^*(a)$.

Conversely, suppose $b \mid a \mid c$. If b and c are both in $\delta(a)$ and if $b \neq b_0$, $c \neq b_0$, then $b_0 \mid a \mid b$ and $b_0 \mid a \mid c$ are false. But then $b_0 \mid b \mid a$ implies by Axiom 3.3 $b_0 \mid a \mid c$; $b \mid b_0 \mid a$ implies by Theorem 19 $b_0 \mid a \mid c$; $\mid b_0, a, b \mid^{\pm}$ implies by Axiom 3.2 $b_0 \mid a \mid c$. Hence there is a contradiction. If, for example, $b_0 = b$, then necessarily $c \subset \delta^*(a)$, and this is a contradiction. If b and c were both in $\delta^*(a)$, then $b_0 \mid a \mid b$ and $b_0 \mid a \mid c$, and Axiom 3.4 is contradicted. Hence a divides E.

This subdivision is unique. For if b is any element of $\delta(a)$, then $\delta^*(a)$ is uniquely determined as the set of all c such that $b \mid a \mid c$. Similarly $\delta(a)$ is unique.

This theorem corresponds to the Jordan theorem as used above (§2.2).

Notation. If a is any element of E, $\delta(a)$ will denote one of the two subsets into which a divides E. If the choice of $\delta(a)$ is fixed, then $\delta^*(a)$ will denote the other of the two sets.

DEFINITION. A subset V of E is *cyclic* if it contains at least one element and, when V contains at least three elements, $\mid a, b, c \mid^{\pm}$ for every triple a, b, c in V.

THEOREM 27. *If V is cyclic and contains at least two elements, then $\delta(a)$ can be chosen uniquely for every a in V so that* (1) $V - a \subset \delta(a)$. *For these choices of $\delta(a)$ we then have*

$$[a \cup \delta^*(a)] \cdot [b \cup \delta^*(b)] = 0 \tag{2}$$

and

$$\delta(a) \supset \delta^*(b) \tag{3}$$

for each pair a, b of distinct elements in V.

Proof. For each a in V take $\delta(a)$ to include some other element b of V. But for any third element c of V, $| a, b, c |^{\pm}$. Hence $c \subset \delta(a)$. Thus $V - a \subset \delta(a)$. The choice of $\delta(a)$ is then unique.

Suppose now a and b distinct and $c \subset \delta^*(b)$. Since $a \subset \delta(b)$, we have by Theorem 26 $c \mid b \mid a$. If c were in $\delta^*(a)$, then similarly $b \mid a \mid c$, and Axiom 1 is contradicted. Hence c is in $\delta(a)$ and (3) follows.

If the left side of (2) is multiplied out, we obtain terms $a \cdot b$, $a \cdot \delta^*(b)$, $b \cdot \delta^*(a)$, $\delta^*(a) \cdot \delta^*(b)$. The first three are void since a and b are distinct and $a \subset \delta(b)$, $b \subset \delta(a)$. The last is void since

$$\delta^*(a) \cdot \delta^*(b) = \delta^*(a) \cdot (\delta(a) \cdot \delta^*(b)) = 0.$$

Hence (2) follows.

Definition. If V is cyclic and contains at least two elements, then $\theta(V)$ will denote the set $\prod_{a \subset V} \delta(a)$, where $\delta(a)$ is chosen in accordance with Theorem 27. If V contains just one element a, $\theta(V)$ will denote one choice of $\delta(a)$. $\theta(V)$ may be regarded as the "interior" of the "region bounded by V".

Theorem 28. *If V is cyclic and contains at least two elements, then the condition $c \subset \theta(V)$ is equivalent to the condition*: $| a, b, c |^{\pm}$ *or* $a \mid c \mid b$ *for every pair* a, b *in* V.

Proof. Suppose c is in $\theta(V)$. Then for every pair a, b, we have $c \subset \delta(a) \cdot \delta(b)$. The condition $b \mid a \mid c$ is equivalent to the condition $c \subset \delta^*(a)$, and this is impossible. Similarly $a \mid b \mid c$ is false. Hence $| a, b, c |^{\pm}$ or $a \mid c \mid b$. Conversely, if $| a, b, c |^{\pm}$ or $b \mid c \mid a$, then c cannot be in $\delta^*(a)$ or $\delta^*(b)$, hence is in $\delta(a) \cdot \delta(b)$; since this holds for every pair, we have $c \subset \theta(V)$.

Definition. If V is a cyclic subset of E and $\theta(V)$ is determined, then $\lambda(V)$ will denote the set $V \cup \theta(V)$. (In case V has only one element, there are two possible choices of $\theta(V)$, hence of $\lambda(V)$.)

Definition. Two sets $\lambda(V_1)$ and $\lambda(V_2)$ will be said to be *adjacent* if V_1 and V_2 have exactly one element a in common and $\delta(a)$ can be so chosen that $\theta(V_1) \subset \delta(a)$, $\theta(V_2) \subset \delta^*(a)$.

Definition. The chordal system E_1 is *isomorphic* to the chordal system E_2 if E_1 can be set into one-to-one correspondence with E_2 in such a way that corresponding triples have the same chordal relations.

Definition. A subset E_1 of E will be said to be *half-parallel* if it is isomorphic to the chordal system of the curve family of parallel lines y = constant in $0 \leqq y < \infty$.

3.3. **Normal subdivisions.** In order to define the normal chordal systems, we shall have to consider an at most countably infinite class of subsets V_α of E. However, the relationships among the V_α being somewhat complicated, it will be necessary to interpret the index α in a special way as a finite sequence of positive integers:

Definition. Let α be a sequence of integers p_i $(i = 1, 2, \cdots, n)$ with $p_i > 0$. Then, for any integer $k > 0$, α, k will denote the sequence $p_1, p_2, \cdots, p_n, k$.

Definition. Let A be a class of finite sequences α of integers > 0. A will be termed *admissible* if the following conditions hold:

(1) A contains the sequence: 1, and no other one-element sequence.

(2) If α, k is in A and $k > 1$, then so also is $\alpha, k - 1$.

(3) If $\alpha, 1$ is in A, then so also is α.

Remark. We can associate with A a generalized "topological tree" with vertices p_α and lines $p_\alpha p_{\alpha,k}$. There are in general infinitely many line segments meeting at a point.

Definition. A subset E_0 of E is *seminormal* if there exists an admissible class A of finite sequences α and a corresponding set of distinct elements a_α of E_0 such that

(1) $E_0 = \delta(a_1) \cup a_1$ for appropriate choice of $\delta(a_1)$;

(2) the sets $V_\alpha = a_\alpha \cup \sum_k a_{\alpha,k}$ are cyclic;

and such that, further, the choices of $\theta(V_\alpha)$ for one-element sets V_α can be so fixed that ($\theta(V_\alpha)$ being now single-valued for all α) for all α:

(3) if α and α, k are in A, then $\lambda(V_\alpha)$ and $\lambda(V_{\alpha,k})$ are adjacent;

(4) $a_\alpha \cup \theta(V_\alpha)$ is half-parallel;

(5) $E_0 = \sum_\alpha \lambda(V_\alpha)$;

(6) for each element a of $V_\alpha - a_\alpha$ there is an element b of $\theta(V_\alpha)$ such that $| a_\alpha, a, b |^{\pm}$.

The V_α are then said to determine a *seminormal subdivision* of E_0.

Definition. E is *normal* if there is an element a of E such that $\delta(a) \cup a$ and $\delta^*(a) \cup a$ are seminormal. If these two sets are seminormally subdivided, then E will be said to be *normally subdivided*.

A normal subdivision of a curve family is illustrated in Figure 3, §3.6.

3.4. **Cross-sections and values of the chordal relations.** We again let F be a regular curve family filling the plane. If G is a subset of F, then G will denote both the corresponding point set which its curves form and the set of elements of CS(F). In each application the particular meaning will be evident from the context.

Theorem 29. *Let the curves C_1 and C_2 of F be joined by a cross-section p_1p_2, with p_1 on C_1, p_2 on C_2. Let S be the set of curves crossing p_1p_2 except at p_1 and p_2. Then S forms an open point set and the condition $C_1 | C_3 | C_2$ is equivalent to the condition that C_3 is in S.*

Proof. Let P be a point on a curve C of S. Let Q be the point at which C meets p_1p_2 and choose an r-neighborhood $U(Q)$ so small that all curves meeting $U(Q)$ cross p_1p_2. By the second condition of regularity there is an r-neighborhood $U(P)$ so small that every curve crossing $U(P)$ crosses $U(Q)$ and hence is in S. Hence S is open.

If $C_1 | C_3 | C_2$, then $\mathfrak{D}(C_1) \subset \mathfrak{D}(C_3) \subset \mathfrak{D}(C_2)$ for proper choices. Hence any curve joining C_1 and C_2 meets C_3 and C_3 is in S.

Conversely, if C_3 is in S, then let C_3 meet p_1p_2 at Q. It follows from con-

sideration of an r-neighborhood $U(Q)$ that C_3 divides p_1p_2 into two segments p_1Q, Qp_2, one of which lies (except for Q) in $\mathfrak{D}(C_3)$, the other in $\mathfrak{D}^*(C_3)$. Hence $C_1 \subset \mathfrak{D}(C_3)$, $C_2 \subset \mathfrak{D}^*(C_3)$ for proper choice of $\mathfrak{D}(C_3)$ and $C_1 \mid C_3 \mid C_2$.

COROLLARY 1. *Let p_1p_3 and p_3p_2 be cross-sections, with p_i on C_i ($i = 1, 2, 3$). Then $p_1p_3p_2$ is a cross-section if and only if $p_1 \subset \mathfrak{D}(C_3)$, $p_2 \subset \mathfrak{D}^*(C_3)$ for proper choice.*

Proof. If $p_1p_3p_2$ is a cross-section, then $C_1 \mid C_3 \mid C_2$, whence $p_1 \subset \mathfrak{D}(C_3)$, $p_2 \subset \mathfrak{D}^*(C_3)$. Conversely, if $p_1 \subset \mathfrak{D}(C_3)$, $p_2 \subset \mathfrak{D}^*(C_3)$, then $p_1p_3p_2$ is a cross-section by Theorem 15.

COROLLARY 2. *If p_1p_2 is a cross-section, then there is a larger cross-section $p_1'p_2'$ containing it in its interior.*

Proof. Let p_1 lie on C_1, p_2 on C_2, and choose $\mathfrak{D}(C_1)$ to include C_2. Choose a cross-section $p_3p_1p_4$ through p_1. Then by Corollary 1 one of p_3 and p_4 is in $\mathfrak{D}^*(C_1)$. Let p_1' denote that one. Then $p_1'p_1p_2$ is a cross-section. Similarly we can find a cross-section $p_1'p_1p_2p_2'$.

DEFINITION. A regular curve-family S filling an open region R is said to be *orientable* if a positive direction can be assigned along each curve of S in such a way that for each point P of R there is an r-neighborhood $U(P)$ in which the curve arcs are similarly directed.

We assume always that an orientable family is properly oriented on each curve.

Whitney has shown ([9], p. 270) that in an orientable regular family of curves there is a function $f(p, t)$ with the properties: for each p in R and any t in $-\infty < t < \infty$ there is a unique point $q = f(p, t)$ lying on the curve C through p; $f(p, t)$ is continuous in both variables (i.e., for each p and t and any $\epsilon > 0$ there is a $\delta > 0$ such that $\rho[f(p, t), f(p', t')] < \epsilon$ if $\rho(p, p') < \delta$ and $|t - t'| < \delta$); $f(p, 0) = p$, and as t increases (decreases) $f(p, t)$ moves continuously in the positive (negative) direction on C.

THEOREM 30. *Under the assumptions of Theorem 29, $S \cup C_1 \cup C_2$ can be mapped homeomorphically on a strip $0 \leqq x \leqq 1$, $-\infty < y < \infty$ in such a way that the curves of $S \cup C_1 \cup C_2$ become the straight lines $x =$ constant.*

Proof. Choose the cross-section $p_1'p_1p_2p_2'$ of Corollary 2 above. Let C_1' and C_2' be the curves through p_1' and p_2' respectively. Let S' be the set of curves crossing $p_1'p_1p_2p_2'$ except at p_1' and p_2'. Then by Theorem 33 S' forms an open set R and S' is a regular family of curves filling R.

Further S' is orientable. For $p_1'p_2'$ divides R into two regions R' and R''. Each curve of S' crosses $p_1'p_2'$ from R' to R''. We assign the positive direction on each curve as from R' to R''. Let P be a point of a curve C of S'. Let C meet $p_1'p_2'$ at Q. By Theorem 8 there is an r-neighborhood $U(P)$ containing the arc PQ of C. The arcs of this neighborhood are then similarly directed. Hence S' is orientable.

Now let $f(p, t)$ be defined as above in S'. Let p vary along the cross-section $p_1'p_1p_2p_2'$, $p = p(\tau)$ in $-1 \leqq \tau \leqq 2$, where we can assume $p(-1) = p_1'$, $p(0) =$

p_1, $p(1) = p_2$, $p(2) = p_2'$. Then $P = f(p(\tau), t)$ gives a homeomorphism T of the strip $-1 < \tau < 2$, $-\infty < t < \infty$ onto S' so that each line $\tau =$ constant becomes a curve of S'. The set $S \cup C_1 \cup C_2$ is thus mapped homeomorphically by T^{-1} on the strip $0 \leqq \tau \leqq 1$, $-\infty < t < \infty$. Set $x = \tau$, $y = t$ and our theorem is established.

COROLLARY. *Under the assumptions of Theorem* 30, *there is a cross-section* q_1q_2 *joining any two points* q_1 *on* C_1 *and* q_2 *on* C_2.

For let the images of q_1 and q_2 under T^{-1} be s_1 and s_2 and draw the straight line segment s_1s_2. $T(s_1s_2)$ is then a cross-section q_1q_2.

3.5. Extended cross-sections tending properly to infinity.

THEOREM 31. *Let C be a curve of F. Then the sets $\delta(C)$ and $\delta^*(C)$ coincide with $\mathfrak{D}(C)$ and $\mathfrak{D}^*(C)$ respectively for suitable choices of δ and $\mathfrak{D}$.*

This follows from Theorem 26 and Definition 1.

THEOREM 32. *Let V be an infinite cyclic subset of $CS(F)$. Then V is countably infinite and the curves C_n of V $(n = 1, 2, \cdots)$ tend uniformly to infinity.*

Proof. By Theorems 31 and 27 we can choose $\mathfrak{D}(C)$ for every C in V so that $\mathfrak{D}(C)\cdot\mathfrak{D}(C') = 0$ for each pair C, C' in V. Each such $\mathfrak{D}(C)$ contains inner points. Hence there are only countably many sets $\mathfrak{D}(C)$ and V is countably infinite. We number its curves as C_n $(n = 1, 2, \cdots)$. If $\{P_n\}$ is a sequence of points with $P_n \subset C_n$ and $\{P_n\}$ has a finite limit point P, then choose a cross-section γ through P and an r-neighborhood of P so small that all curves crossing it cross γ. It follows that there are three curves C_r, C_s, C_t of the sequence C_n crossing γ. By Theorem 29 $C_r \mid C_s \mid C_t$, $C_s \mid C_t \mid C_r$, or $C_t \mid C_r \mid C_s$, and this is a contradiction. Hence the curves C_n tend uniformly to infinity.

THEOREM 33. *Let C_0 be a curve of F, K a circle of radius R, center at P on C_0, $\mathfrak{D}(C_0)$ chosen fixed. Then there is a curve of F lying in $\mathfrak{D}(C_0)$ and not meeting K.*

Proof. For each curve C in $\mathfrak{D}(C_0)$ choose $\mathfrak{D}(C) \subset \mathfrak{D}(C_0)$. Set $d(C) =$ distance from P to C. Then $C'' \mid C' \mid C_0$ implies $d(C') < d(C'')$. Suppose every C in $\mathfrak{D}(C_0)$ meets K. Then $d(C)$ has a positive finite least upper bound r. We can then choose a sequence C_n of curves in $\mathfrak{D}(C_0)$ such that $d(C_n)$ increases monotonely to r. Take the point Q_n in $C_n\cdot K$. Then the Q_n have a limit point Q on a curve C_Q. Take a cross-section γ through Q. Assume the C_n restricted to a subsequence (which we again call C_n) such that the intersection points of the C_n with γ approach Q monotonely along γ. Hence $C_{n_1} \mid C_{n_2} \mid C_Q$ for $n_2 > n_1$ and both sufficiently large.

Further $C_0 \mid C_{n_2} \mid C_Q$. For $C_{n_2} \mid C_0 \mid C_Q$ is impossible, since $C_{n_2} \subset \mathfrak{D}(C_0)$, $C_Q \subset \mathfrak{D}(C_0)$. $C_0 \mid C_Q \mid C_{n_2}$ implies by Axiom 3.3 $C_0 \mid C_{n_2} \mid C_{n_1}$. Hence $d(C_{n_1}) > d(C_{n_2})$, and this is impossible. $\mid C_0, C_{n_2}, C_Q \mid^{\pm}$ implies by Axiom 3.2 $C_0 \mid C_{n_2} \mid C_{n_1}$, and this is also impossible. Thus $C_0 \mid C_{n_2} \mid C_Q$. But that implies $d(C_Q) \geqq r$. Choose a curve C' in $\mathfrak{D}(C_Q)$, whence $C' \mid C_Q \mid C_0$ and $d(C') > d(C_Q) > r$. This is a contradiction, and the theorem follows.

THEOREM 34. *Let C_0 be a curve of F, $\mathfrak{D}(C_0)$ chosen fixed, P a point on C_0. Let M denote the set of all points Q such that there is a cross-section PQ. Let H denote $M \cdot \mathfrak{D}(C_0)$. Then $M \cup C_0$ and H are non-void open simply-connected sets and are unions of curves of F. The boundary of H includes C_0 and is a cyclic subset of F, that of $M \cup C_0$ is either void or a cyclic subset of F.*

Proof. By the corollary to Theorem 30, $M \cup C_0$ is a union of curves of F. Every point of $M \cup C_0$ is joined to P by an arc in $M \cup C_0$, hence $M \cup C_0$ is connected. If Q is in M, choose a cross-section PQQ_1, by the Corollary 2 to Theorem 29, and then an r-neighborhood $U(Q)$ so small that every curve meeting it meets PQQ_1. Hence $U(Q) \subset M$. If $Q \subset C_0$, then any r-neighborhood $U(Q)$ is in $M \cup C_0$. Hence $M \cup C_0$ is open.

If Γ is a closed curve lying wholly in $M \cup C_0$, then no point Q interior to Γ can lie without $M \cup C_0$, for then the curve C_Q would intersect Γ, and this is a contradiction. Hence $M \cup C_0$ is simply-connected.

Let Q be a boundary point of $M \cup C_0$, Q' any other point of the curve C_Q. Then the arc QQ' lies in an r-neighborhood, by Theorem 8. Hence Q' is a boundary point. The boundary of $M \cup C_0$ is thus either void or a union of curves of F. If there are at least three boundary curves, let J_1, J_2, J_3 be such. $J_2 \mid J_1 \mid J_3$ implies $J_2 \subset \mathfrak{D}(J_1)$, $J_3 \subset \mathfrak{D}^*(J_1)$, and the connectedness of $M \cup C_0$ is contradicted. Similarly $J_1 \mid J_2 \mid J_3$ and $J_1 \mid J_3 \mid J_2$ are excluded. Therefore $\mid J_1, J_2, J_3 \mid^{\pm}$ and the boundary is cyclic.

By Jordan's theorem, C_0 divides $M \cup C_0$ into two open simply-connected sets, one of which is H, and H is necessarily a union of curves of F. Furthermore the boundary of H includes C_0 and, by the same argument as above, is a cyclic subset of F.

Remark. The set H need not have the structure of the parallel lines and may indeed be very complicated.

THEOREM 35. *Under the assumptions of Theorem 34, for every circle K of center P and radius R there is a curve C_0' in H not meeting K.*

Proof. Since $M \cup C_0$ is an open simply-connected set, it can be mapped on the entire plane by a homeomorphism T; the curves of F in $M \cup C_0$ are transformed onto a curve family filling the plane. Under T, H becomes a set $\mathfrak{D}(T(C_0))$, K becomes a closed curve (hence a bounded set). It follows from Theorem 33 that there is a curve $T(C')$ of $T(F)$ in $\mathfrak{D}(T(C_0))$ and not meeting $T(K)$. The inverse C_0' of $T(C_0')$ then lies in H and does not meet K.

DEFINITION. An *extended cross-section* will mean a curve Γ which meets each curve of F at most once. Γ will further be said to *tend properly to infinity* in a given direction on Γ if it tends to infinity in that direction and in such a way that the curves meeting it tend uniformly to infinity with their intersection points with Γ.

THEOREM 36. *Let C_0 be a curve of F, $\mathfrak{D}(C_0)$ chosen fixed, P a point on C_0. Then there is an extended cross-section Γ from P to ∞ in $\mathfrak{D}(C_0)$ and tending properly to ∞.*

Proof. Let $K_n(P)$ denote a circle with center P and radius n $(n = 1, 2, \cdots)$. For every Q in $\mathfrak{D}(C_0)$ let $K_n(Q)$ denote a circle with center Q and radius so large that $K_n(P)$ lies within $K_n(Q)$. Choose $\mathfrak{D}(C) \subset \mathfrak{D}(C_0)$ for every C in $\mathfrak{D}(C_0)$.

Now, in accordance with Theorem 35, we choose C_1 in $\mathfrak{D}(C_0)$ outside of $K_1(P)$ and Q_1 a point on C_1. Next, using Theorem 35, choose C_2 in $\mathfrak{D}(C_1)$ outside of $K_2(Q_1)$ and Q_2 on C_2. Proceeding in this way, we obtain a sequence of curves $C_0, C_1, C_2, \cdots, C_n, \cdots$ so that $\mathfrak{D}(C_0) \supset \mathfrak{D}(C_1) \supset \cdots \supset \mathfrak{D}(C_n) \supset \cdots$ and a sequence of points $P, Q_1, Q_2, \cdots, Q_n, \cdots$ so that there are cross-sections $PQ_1, Q_1Q_2, \cdots, Q_nQ_{n+1}, \cdots$. By Corollary 2 to Theorem 29, $\Gamma_n = PQ_1Q_2 \cdots Q_n$ is a cross-section. Thus, as $n \to \infty$, Γ_n approaches a limiting half-open curve Γ which is an extended cross-section. Further, since C_n, and hence $\mathfrak{D}(C_n)$, fails to meet $K_n(P)$, Γ tends properly to ∞.

3.6. Normal subdivision of CS(F).

THEOREM 37. *Let Γ be a half-open curve from C_0 to ∞ which forms an extended cross-section. Then the set S of curves of F crossing Γ forms a half-parallel subset of $CS(F)$. The set $S - C_0$ is an open simply-connected set whose boundary is a cyclic subset V of $CS(F)$. Further $S - C_0 = \theta(V)$.*

Proof. Let Γ be given by $x = f(t)$, $y = g(t)$ in $0 \leqq t < \infty$. Let C_t be the curve meeting Γ at P_t: $x = f(t)$, $y = g(t)$. From Theorem 29 we conclude that $t_1 < t_2 < t_3$ implies $C_{t_1} \mid C_{t_2} \mid C_{t_3}$. If D_t is the line $y = t$ of the xy-plane, then $\varphi(C_t) = D_t$ gives an isomorphism of S onto the set of lines D_t in $0 \leqq t < \infty$. Hence S is half-parallel.

$S - C_0$ is connected since all its curves cross Γ. It is open, by Theorem 29, since it is a union of open sets. A closed curve γ in $S - C_0$ cannot include a boundary point Q of $S - C_0$ in its interior, for then the curve C_Q would cross γ. Hence $S - C_0$ is simply-connected. The fact that the boundary of $S - C_0$ forms a cyclic subset V of F follows as in the proof of Theorem 34. V includes C_0.

By Theorem 32, V is at most countably infinite and we number its curves as J_n $(n = 1, 2, \cdots)$ with $J_1 = C_0$. Choose $\mathfrak{D}(J_n) \supset V - J_n$, then $\theta(V) = \prod_1^\infty \mathfrak{D}(J_n)$. If C_t is a curve of S, then for any J_{n_1}, J_{n_2} of V, $C_t \mid J_{n_1} \mid J_{n_2}$ or $C_t \mid J_{n_2} \mid J_{n_1}$ would contradict the connectedness of S. Hence $J_{n_1} \mid C_t \mid J_{n_2}$ or $\mid J_{n_1}, J_{n_2}, C_t \mid^{\pm}$. Hence, by Theorem 28, $C_t \subset \theta(V)$. Conversely, if C is a curve of $\theta(V)$, then C is in S. For suppose C is not in S. Then, since by Theorem 32 the J_n tend uniformly to infinity, a point on C has a positive distance from the set $\sum_n J_n$. There must thus exist a line segment PQ joining the point P on C to the point Q on some J_{n_0} and otherwise containing no points of C or of $\sum_n J_n$. C lies in $\mathfrak{D}(J_{n_0})$, as does S. Since J_{n_0} bounds S, PQ must contain points of S, and there must be a boundary point of S other than Q along PQ. This is impossible. Hence $S - C_0 = \theta(V)$.

THEOREM 38. *Let F be a regular family filling the plane. Then $CS(F)$ is normal.*

Preliminary outline of proof. We take an extended cross-section Γ as in Theorem 36. From each boundary curve of S we take new extended cross-sections to infinity. Repeating the process indefinitely, we divide the family up into non-overlapping half-parallel sets. The only difficulty is to ensure that the whole family is exhausted by the process. This we do by covering the plane by a countable number of r-neighborhoods and making sure that these are all included as the process expands.

Proof of theorem. Take a fixed point O of the plane on a curve C_1 of F. Choose $\mathfrak{D}(C_1)$ fixed. We shall then carry out a seminormal subdivision of $C_1 \cup \mathfrak{D}(C_1)$ corresponding to an admissible class A of finite sequences of positive integers. We shall let A_n denote the subclass of A whose sequences have n terms $(n = 1, 2, \cdots)$.

For every curve C of $\mathfrak{D}(C_1)$ choose $\mathfrak{D}(C) \subset \mathfrak{D}(C_1)$. By Theorem 36, for each point Q on C we can find an extended cross-section $H(Q)$ tending properly to ∞ in $\mathfrak{D}(C)$ and another extended cross-section $H^*(Q)$ tending properly to ∞ in $\mathfrak{D}^*(C)$. By Theorem 15, $H(Q)$ plus $H^*(Q)$ forms an open extended cross-section $\Gamma(Q)$. The set $S(Q)$ of curves crossing $\Gamma(Q)$ is by Theorem 29 an open set. Hence we can find an r-neighborhood $U(Q)$ lying in $S(Q)$.

Assume now a fixed choice of $U(Q)$ for each point Q of $\mathfrak{D}(C_1) \cup C_1$. Let K_N be the circle center O and radius N $(N = 1, 2, \cdots)$ plus its interior. Then by the Heine-Borel Theorem a finite number of the r-neighborhoods cover the set $K_N \cdot (\mathfrak{D}(C_1) \cup C_1)$. We can thus choose a sequence $U(Q_i)$ $(i = 1, 2, \cdots)$ such that for r sufficiently large the set $K_N \cdot (\mathfrak{D}(C_1) \cup C_1)$ is included in $\sum_{i=1}^{r} U(Q_i)$. We can assume Q_1 is O. If Q is any point of $\mathfrak{D}(C_1) \cup C_1$, we shall denote by $U(Q_i, Q)$ the $U(Q_i)$ of lowest index i containing Q. We shall then denote by $\Gamma(Q_i, Q)$ the half-open curve on $\Gamma(Q_i)$ consisting of $\Gamma(Q_i) \cdot (\mathfrak{D}(C_Q) \cup C_Q)$. Since Q is in $U(Q_i)$, C_Q must necessarily meet $\Gamma(Q_i)$ and $\Gamma(Q_i, Q)$ consists of the part of $\Gamma(Q_i)$ joining the intersection point to ∞ in $\mathfrak{D}(C_Q)$. $S(Q_i, Q)$ will then denote the set of curves crossing $\Gamma(Q_i, Q)$. $S(Q_i, Q)$ is then half-parallel, by Theorem 37.

The class A_1 will now consist of the single one-element sequence: 1. C_1 will be the corresponding element of $CS(F)$.

Suppose now that the classes $A_1, A_2, \cdots, A_r$ have been defined and also the curves C_α for α in $A_1, A_2, \cdots, A_r$. Then for α in A_r choose a point Q_α on C_α at minimum distance from O. The boundary of the set $S(Q_i, Q_\alpha)$ is then by Theorem 37 a cyclic set V_α, at most countably infinite, and its curves can be given as $C_\alpha, C_{\alpha,1}, C_{\alpha,2}, \cdots$. The indices $\alpha, 1; \alpha, 2; \cdots$, with α varying over A_r, then form the class A_{r+1}. The class A of all A_n is now defined and admissible, and the curves C_α are defined for all α in A. (See Figure 3.)

Condition (2) is now fulfilled. If we determine $\theta(V_\alpha)$ in the case of a one-

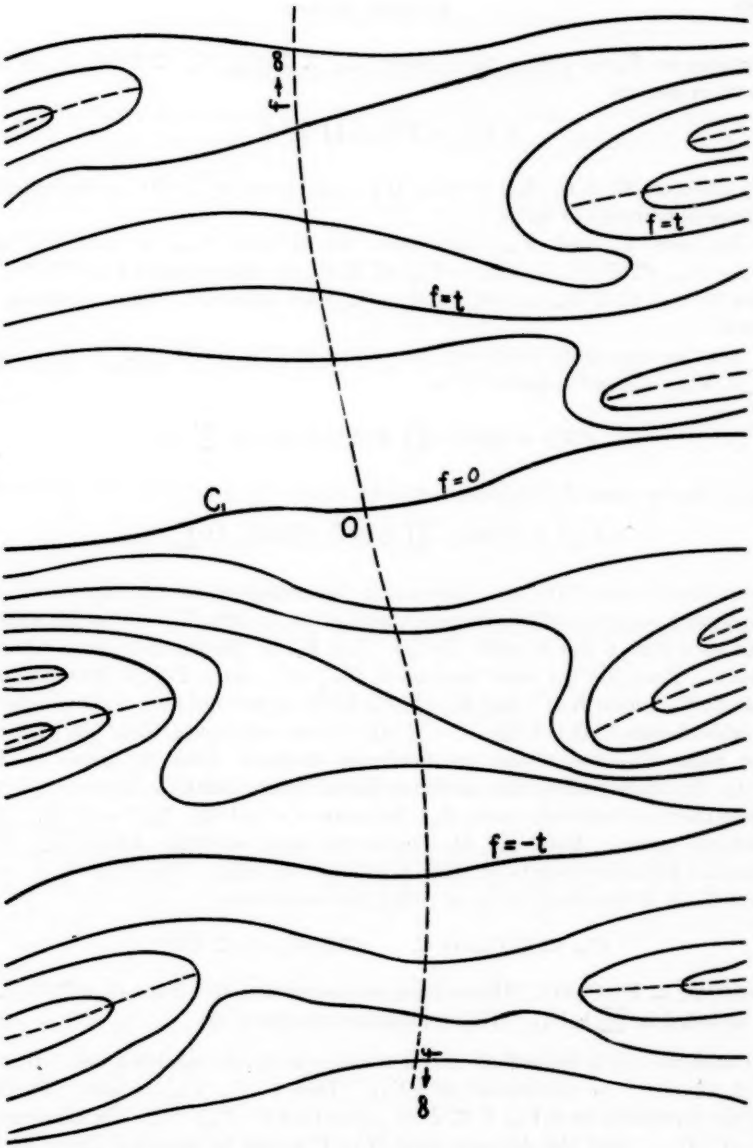

FIG. 3. NORMAL SUBDIVISION OF A CURVE-FAMILY

Each cross-section (dotted line) joins a finite point to ∞ and each determines a set of curves homeomorphic with the parallel lines of a half-plane. The assignment of values for a level-curve function $f(x, y)$ is suggested on the figure.

element set V_α as $\mathfrak{D}(C_\alpha)$, then, since $C_{\alpha,k} \subset \mathfrak{D}(C_\alpha)$, $C_\alpha \subset \mathfrak{D}^*(C_{\alpha,k})$, we shall have in general

$$\theta(V_\alpha) = \mathfrak{D}(C_\alpha) \cdot \prod_k \mathfrak{D}^*(C_{\alpha,k}).$$

By Theorem 37, $S(Q_i, Q_\alpha) = \theta(V_\alpha) \cup C_\alpha$, and hence $\theta(V_\alpha) \cup C_\alpha$ is half-parallel. Hence condition (4) holds.

The sets V_α and $V_{\alpha,k}$ have only the element $C_{\alpha,k}$ in common, since $V_\alpha - C_{\alpha,k} \subset \mathfrak{D}^*(C_{\alpha,k})$, $V_{\alpha,k} - C_{\alpha,k} \subset \mathfrak{D}(C_{\alpha,k})$. Since also $\theta(V_\alpha) \subset \mathfrak{D}^*(C_{\alpha,k})$ and $\theta(V_{\alpha,k}) \subset \mathfrak{D}(C_{\alpha,k})$, $\lambda(V_\alpha)$ and $\lambda(V_{\alpha,k})$ are adjacent. Thus condition (3) holds.

Next we can verify condition (5). Suppose C is in $\mathfrak{D}(C_1) \cup C_1$, but not in $\sum_\alpha \lambda(V_\alpha)$. Then C cannot lie in

$$\lambda(V_1) = [\mathfrak{D}(C_1) \cdot \prod_k \mathfrak{D}^*(C_{1,k})] \cup C_1 \cup \sum_k C_{1,k},$$

hence lies in some $\mathfrak{D}(C_{1,k})$; cannot lie in

$$\lambda(V_{1,k}) = [\mathfrak{D}(C_{1,k}) \cdot \prod_l \mathfrak{D}^*(C_{1,k,l})] \cup C_{1,k} \cup \sum_l C_{1,k,l},$$

hence lies in some $\mathfrak{D}(C_{1,k,l})$. Repeating this argument, we find that there is an expanding sequence of sequences α^n, with α^n in A_n and $\alpha^{n+1} = \alpha^n, k$ for some k, and such that C lies in each $\mathfrak{D}(C_{\alpha^n})$. Let K_N be the smallest circle which C meets. Then $\mathfrak{D}(C_{\alpha^n})$ must also meet K_N, and, since $\mathfrak{D}(C_{\alpha^n})$ does not contain O, C_{α^n} meets K_N. But $K_N \cdot (\mathfrak{D}(C_1) \cup C_1)$ is covered by a finite number of neighborhoods $U(Q_1)$, $U(Q_2)$, $\cdots$, $U(Q_s)$ of the sequence $U(Q_i)$. On each C_{α^n} the point Q_{α^n} was chosen at minimum distance from O, hence in K_N. $U(Q_i, Q_{\alpha^n})$ was chosen with minimum index i to contain Q_{α^n}, hence $1 \leqq i \leqq s$. Since there are infinitely many Q_{α^n}, for some $n < p$ $U(Q_i, Q_{\alpha^n}) \equiv U(Q_i, Q_{\alpha^p})$, with the same i. But $U(Q_i, Q_{\alpha^n})$ lies in the open set $S(Q_i)$, while $U(Q_i, Q_{\alpha^p})$ contains boundary points or other points not in $S(Q_i)$. For, if $p = n + 1$, then C_{α^p} is a boundary curve of $S(Q_i)$, and otherwise

$$C_{\alpha^p} \subset \mathfrak{D}(C_{\alpha^{p-1}}) \subset \cdots \subset \mathfrak{D}(C_{\alpha^{n+1}}) \subset \mathfrak{D}(C_{\alpha^n})$$

and $S(Q_i) \subset \mathfrak{D}^*(C_{\alpha^{n+1}})$. Hence there is a contradiction. Thus all of $\mathfrak{D}(C_1) \cup C_1$ is included in $\sum_\alpha \lambda(V_\alpha)$. This establishes condition (5).

Condition (1) is immediate and it remains to verify condition (6). But for each $C_{\alpha,k}$ let C be an element of $\theta(V_\alpha)$. Then $C \mid C_\alpha \mid C_{\alpha,k}$ is false. Further, by the expression for $\theta(V_\alpha)$, $C \subset \mathfrak{D}^*(C_{\alpha,k})$ and not $C \mid C_{\alpha,k} \mid C_\alpha$. If, for every C, $C_\alpha \mid C \mid C_{\alpha,k}$, then the distance from O to C would be bounded above by the distance from O to $C_{\alpha,k}$. This is impossible since $\Gamma(Q_\alpha)$ tends properly to infinity. Hence $\mid C_\alpha, C, C_{\alpha,k} \mid^{\pm}$ for some C in $\theta(V_\alpha)$.

We thus conclude that $\mathfrak{D}(C_1) \cup C_1$ is seminormal and by the same reasoning that $\mathfrak{D}^*(C_1) \cup C_1$ is seminormal. Hence $\mathrm{CS}(F)$ is normal.

4. Existence of a level-curve function

4.1. **Further information about the normal subdivision.** We first assume an abstract normal chordal system E with a seminormal subdivision of the set $E_0 = \delta(c_1) \cup c_1$ as defined above.

THEOREM 39. $\lambda(V_\alpha)\cdot\lambda(V_{\alpha'}) = 0$ *unless* $\alpha = \alpha', k$ *for some* k *or* $\alpha' = \alpha, k$ *for some* k.

Proof. Since $c_{1,k} \subset \delta(c_1)$, we must have

$$\theta(V_1) = \delta(c_1)\cdot\prod_k \delta^*(c_{1,k})$$

for proper choices of the $\delta^*(c_{1,k})$. Since $\lambda(V_1)$ and $\lambda(V_{1,k})$ are adjacent, we must have

$$\theta(V_{1,k}) = \delta(c_{1,k})\cdot\prod_l \delta^*(c_{1,k,l}).$$

Repeating this reasoning, we conclude that in general

$$\theta(V_\alpha) = \delta(c_\alpha)\cdot\prod_k \delta^*(c_{\alpha,k})$$

for suitable fixed choices of $\delta(c_\alpha)$.

Hence $\lambda(V_\alpha) = c_\alpha \cup \sum_k c_{\alpha,k} \cup [\delta(c_\alpha)\cdot\prod_k \delta^*(c_{\alpha,k})]$ by Theorem 27. Accordingly we then have

$$V_{\alpha,k} \subset c_{\alpha,k} \cup \delta(c_{\alpha,k}) \subset \delta(c_\alpha)$$

and for $k \neq k'$

$$\lambda(V_{\alpha,k})\cdot\lambda(V_{\alpha,k'}) \subset [c_{\alpha,k} \cup \delta(c_{\alpha,k})]\cdot[c_{\alpha,k'} \cup \delta(c_{\alpha,k'})] = 0$$

from Theorem 27.

Now for any V_α and $V_{\alpha'}$ with $\alpha \neq \alpha'$, let α_0 be the largest sequence such that $\alpha = \alpha_0, p, \cdots, \alpha' = \alpha_0, p', \cdots$, with $p \neq p'$. The case $\alpha = \alpha_0, p, \alpha' = \alpha_0, p'$ we have just treated. If $\alpha = \alpha_0, p_1, p_2, \cdots, p_r$, with $r > 1$, and $\alpha' = \alpha_0, p_1', p_2', \cdots, p_{r'}'$, with $r' \geqq 1$, then

$$\lambda(V_\alpha) \subset \delta(c_{\alpha_0,p_1,\cdots,p_r}) \subset \cdots \subset \delta(c_{\alpha_0,p_1})$$

while $\lambda(V_{\alpha'}) \subset \delta(c_{\alpha_0,p_1'}) \cup c_{\alpha_0,p_1'}$, whence $\lambda(V_\alpha)\cdot\lambda(V_{\alpha'}) = 0$ by Theorem 27. If $\alpha = \alpha_0, p_1, p_2, \cdots, p_r$, with $r > 1$, and $\alpha' = \alpha_0$, then $\lambda(V_{\alpha_0}) \subset \delta^*(c_{\alpha_0,p_1'})$ and again $\lambda(V_\alpha)\cdot\lambda(V_{\alpha'}) = 0$. Hence the theorem follows.

Now consider a fixed set $\lambda(V_\alpha)$. The set $c_\alpha \cup \theta(V_\alpha)$ is half-parallel. Hence we can represent its elements as d_t^α for $0 \leqq t < \infty$, whereby $d_{t_1}^\alpha \mid d_{t_2}^\alpha \mid d_{t_3}^\alpha$ is equivalent to the condition that t_2 is between t_1 and t_3. d_0^α must necessarily be c_α, since otherwise $\theta(V_\alpha)$ could not be in $\delta(c_\alpha)$.

Notation. Let $c_{\alpha,k}$ be any element of $V_\alpha - c_\alpha$. By the condition (6) of normality there exists an element d_t^α of $\theta(V_\alpha)$ such that $\mid c_\alpha, c_{\alpha,k}, d_t^\alpha \mid^\pm$. We shall let $B(c_{\alpha,k})$ be the set of all d_t^α such that $\mid c_\alpha, c_{\alpha,k}, d_t^\alpha \mid^\pm$.

THEOREM 40. *For each element $c_{\alpha,k}$ of $V_\alpha - c_\alpha$ there is a unique number $t_{\alpha,k} \geqq 0$ such that either* $B(c_{\alpha,k}) = \sum_{t \geqq t_{\alpha,k}} d_t^\alpha$ *or else* $B(c_{\alpha,k}) = \sum_{t > t_{\alpha,k}} d_t^\alpha$.

Proof. If $d_{t_1}^\alpha \subset B(c_{\alpha,k})$ and $t_2 > t_1$, then $c_\alpha \mid d_{t_1}^\alpha \mid d_{t_2}^\alpha$. Hence, by Axiom 3.2, $\mid c_\alpha, c_{\alpha,k}, d_{t_2}^\alpha \mid^\pm$ and $d_{t_2}^\alpha \subset B(c_{\alpha,k})$. Thus $B(c_{\alpha,k})$ has one of the two forms indicated.

4.2. **Application to a curve family.** We return to the family F, which we assume normally subdivided as above. Denote the curves of the half-parallel set $\theta(V_\alpha) \cup C_\alpha$ by D_t^α in $0 \leqq t < \infty$ (corresponding to the d_t^α of the preceding section).

THEOREM 41. *Let PQ be a cross-section joining a point P on a curve $D_{t_0}^\alpha$ to a point Q on a curve $C_{\alpha,k}$.*

(A) *If $D_{t_{\alpha,k}}^\alpha \subset B(C_{\alpha,k})$, then $t_0 < t_{\alpha,k}$ and the set of curves crossing PQ is given by all D_t^α with $t_0 \leqq t < t_{\alpha,k}$ and the curve $C_{\alpha,k}$.*

(B) *If $D_{t_{\alpha,k}}^\alpha \not\subset B(C_{\alpha,k})$, then $t_0 > t_{\alpha,k}$ and the set of curves crossing PQ is given by all D_t^α with $t_{\alpha,k} < t \leqq t_0$ and the curve $C_{\alpha,k}$.*

Proof. PQ must lie in $\theta(V_\alpha)$ except for Q, for if PQ meets a boundary curve at P_1 it must from P_1 on lie outside $\theta(V_\alpha)$. For any point Q_1 between P and Q on PQ the arc PQ_1 is a cross-section in $\theta(V_\alpha)$. Since $\theta(V_\alpha)$ is half-parallel, PQ_1 meets precisely the D_t^α of an interval[14] $[t_0, t_1]$. Thus PQ meets precisely those D_t^α of a half-open interval $[t_0, t')$.

Case (A). We have then $\mid C_\alpha, C_{\alpha,k}, D_{t_{\alpha,k}}^\alpha \mid^\pm$ and we must show $t_0 < t' = t_{\alpha,k}$. But we can then choose $\mathfrak{D}(D_{t_{\alpha,k}}^\alpha)$ to include $C_{\alpha,k}$ and C_α. Suppose PQ were to cross $D_{t_{\alpha,k}}^\alpha$ at a point P_1. Then P_1Q must lie in $\mathfrak{D}(D_{t_{\alpha,k}}^\alpha) \cup D_{t_{\alpha,k}}^\alpha$. But for a curve $D_{t_2}^\alpha$ meeting P_1Q between P_1 and Q we then have $t_2 < t_{\alpha,k}$. Further $C_{\alpha,k} \mid D_{t_2}^\alpha \mid D_{t_{\alpha,k}}^\alpha$, $C_\alpha \mid D_{t_2}^\alpha \mid D_{t_{\alpha,k}}^\alpha$, and also $C_\alpha \mid D_{t_2}^\alpha \mid C_{\alpha,k}$, since $D_{t_2}^\alpha \not\subset B(C_{\alpha,k})$. This contradicts Axiom 3.4. Hence PQ fails to cross $D_{t_{\alpha,k}}^\alpha$. But PQ must meet curves D_t^α with $t < t_{\alpha,k}$, since C_α and $C_{\alpha,k}$ are in $\mathfrak{D}(D_{t_{\alpha,k}}^\alpha)$. Further $D_{t_1}^\alpha \mid D_{t_2}^\alpha \mid C_{\alpha,k}$ for $t_1 < t_2 < t_{\alpha,k}$, as follows from $C_\alpha \mid D_{t_1}^\alpha \mid D_{t_2}^\alpha$, $C_\alpha \mid D_{t_2}^\alpha \mid C_{\alpha,k}$ and Axiom 3.3. Thus PQ must meet all D_t^α for t sufficiently near $t_{\alpha,k}$ and $< t_{\alpha,k}$. This is possible only if $t_0 < t' = t_{\alpha,k}$.

Case (B). We must show $t_{\alpha,k} = t' < t_0$. Suppose $t_{\alpha,k} \neq 0$. We have then, by Theorem 28, $C_\alpha \mid D_{t_{\alpha,k}}^\alpha \mid C_{\alpha,k}$. If PQ were to cross $D_{t_{\alpha,k}}^\alpha$ at P_1, then take a curve $D_{t_2}^\alpha$ meeting P_1Q between P_1 and Q. We must have $t_2 > t_{\alpha,k}$. Hence $C_\alpha \mid D_{t_{\alpha,k}}^\alpha \mid D_{t_2}^\alpha$. But also $D_{t_{\alpha,k}}^\alpha \mid D_{t_2}^\alpha \mid C_{\alpha,k}$ by Theorem 29. Hence, by Axiom 3.3, $C_\alpha \mid D_{t_2}^\alpha \mid C_{\alpha,k}$. This contradicts Theorem 40. Thus PQ fails to cross $D_{t_{\alpha,k}}^\alpha$. But PQ must meet curves D_t^α with $t > t_{\alpha,k}$, since $C_\alpha \mid D_{t_{\alpha,k}}^\alpha \mid C_{\alpha,k}$. Further, for $t_{\alpha,k} < t_1 < t_2$, $\mid C_{\alpha,k}, C_\alpha, D_{t_1}^\alpha \mid^\pm$ and $C_\alpha \mid D_{t_1} \mid D_{t_2}$. Hence by Axiom 3.2, $C_{\alpha,k} \mid D_{t_1} \mid D_{t_2}$. Thus PQ must meet all D_t^α for t sufficiently near $t_{\alpha,k}$ and $> t_{\alpha,k}$. This is possible only if $t_{\alpha,k} = t' < t_0$.

[14] We use the following convention on intervals: $[a, b]$, $(a, b]$, $[a, b)$, (a, b) mean respectively (whether $a < b$ or $b < a$) the closed interval from a to b, the closed interval minus a, the closed interval minus b, the open interval.

If $t_{\alpha,k} = 0$, then $D^{\alpha}_{t_{\alpha,k}} = C_\alpha$ and $| C_\alpha, C_{\alpha,k}, D^\alpha_t |^{\pm}$ for all $t > 0$. PQ cannot meet C_α since $PQ \subset \theta(V_\alpha) \cup C_{\alpha,k}$. PQ must meet curves D^α_t with $t > 0$, and the same reasoning as in the preceding paragraph shows that PQ meets all D^α_t for t sufficiently small and > 0. Hence $t_{\alpha,k} = t' < t_0$.

COROLLARY. *PQ can be parametrized by t in the interval $[t_{\alpha,k}, t_0]$ so that to each value $t \neq t_{\alpha,k}$ corresponds the point at which D^α_t crosses PQ.*

The corollary is true since $t_1 < t_2 < t_3$ implies that $D^\alpha_{t_1} | D^\alpha_{t_2} | D^\alpha_{t_3}$ and hence by Theorem 29 that $D^\alpha_{t_2}$ crosses PQ between $D^\alpha_{t_1}$ and $D^\alpha_{t_3}$.

THEOREM 42. *Let F be a regular family of curves filling the plane. Then there exists a function $f(x, y)$ with the properties:*

(1) *$f(x, y)$ is defined and continuous for all (x, y);*

(2) *for every real number c the locus $f(x, y) = c$ consists of an at most countably infinite set of curves of F;*

(3) *in every neighborhood of any point (x_0, y_0) there are points (x, y) for which $f(x, y) > f(x_0, y_0)$ and points (x, y) for which $f(x, y) < f(x_0, y_0)$.*

Proof. Let the subset $C_1 \cup \mathfrak{D}(C_1)$ of F be subdivided seminormally as above by the V_α and let $C_1 \cup \mathfrak{D}^*(C_1)$ be seminormally subdivided by sets V^*_α for α in A^*. In each $\lambda(V_\alpha) = C_\alpha \cup \sum_k C_{\alpha,k} \cup \sum_{t>0} D^\alpha_t$ we define a function $f_\alpha(x, y)$ thus: (a) for any point $Q:(x, y)$ on a curve D^α_t we set $f_\alpha(x, y) = t = t(x, y)$; (b) for any point $Q:(x, y)$ on a curve $C_{\alpha,k}$ we set $f_\alpha(x, y) = t_{\alpha,k}$.

$f_\alpha(x, y)$ is then continuous in the closed set $\lambda(V_\alpha)$. For in case (a) we take a cross-section γ in $C_\alpha \cup \theta(V_\alpha)$ and containing Q. This can be parametrized by t so that D^α_t meets γ at the point corresponding to t, as in the above corollary. If we then map the set of curves crossing γ onto a strip $0 \leqq x' \leqq 1$, $-\infty < y' < \infty$, as in Theorem 30, then $f_\alpha(x, y)$ becomes the function $\bar{f}_\alpha(x', y') \equiv x'$. Hence $f_\alpha(x, y)$ is continuous at each Q in case (a). In case (b) we choose a cross-section PQ joining Q to a point P of $\theta(V_\alpha)$. Then, by the above corollary, the same reasoning applies to show that $f_\alpha(x, y)$ is continuous at Q.

We define the functions $f^*_\alpha(x, y)$ in the same way in the $\lambda(V^*_\alpha)$, and they are continuous there. We note further that $f_\alpha(x, y) = 0$ for (x, y) on C_α, $f^*_\alpha(x, y) = 0$ for (x, y) on C^*_α.

Further let $\epsilon_{\alpha,k} = +1$ or -1 according as $D^\alpha_{t_{\alpha,k}} \subset B(C_{\alpha,k})$ or $D^\alpha_{t_{\alpha,k}} \not\subset B(C_{\alpha,k})$. If $\alpha, k = 1, k_2, \cdots, k_n, k$, then set

$$\delta_{\alpha,k} = \epsilon_{1,k_2} \cdot \epsilon_{1,k_2,k_3} \cdot \cdots \cdot \epsilon_\alpha \cdot \epsilon_{\alpha,k}.$$

We now set $f(x, y) = f_1(x, y)$ in $\lambda(V_1)$. Suppose we have defined $f(x, y)$ for (x, y) in $\lambda(V_\alpha)$ for all sequences α of at most n elements. Let $\alpha, k = 1, k_2, k_3, \cdots, k_n, k_{n+1}$. Let $f(C_{\alpha,k})$ be the value already assigned on $C_{\alpha,k}$ as boundary of $\lambda(V_\alpha)$. Then, for (x, y) in $\lambda(V_{\alpha,k})$ we set

$$f(x, y) = f(C_{\alpha,k}) + \delta_{\alpha,k} f_{\alpha,k}(x, y).$$

By Theorem 39, $f(x, y)$ is then uniquely defined. On $C_{\alpha,k}$ there is no difficulty since $f_{\alpha,k}(x, y) = 0$ on $C_{\alpha,k}$. Hence $f(x, y)$ becomes a continuous function in all of $C_0 \cup \mathfrak{D}(C_0)$.

Similarly we construct the function $f^*(x, y)$ in $C_0 \cup \mathfrak{D}^*(C_0)$. Finally set $f(x, y) = -f^*(x, y)$ in $C_0 \cup \mathfrak{D}^*(C_0)$. Then $f(x, y)$ is defined, single-valued, and continuous everywhere.

In the neighborhood of a point (x_0, y_0) of $\theta(V_\alpha)$, $f(x, y)$ takes on values both greater and less than $f(x_0, y_0)$. For along a cross-section through (x_0, y_0) $f_\alpha(x, y)$ varies monotonely, and the same holds for $f(x, y)$. The same holds for points of $\theta(V_\alpha^*)$.

For a point Q: (x_0, y_0) on a curve $C_{\alpha,k}$, we first draw a cross-section QP joining Q to a point P of $\theta(V_\alpha)$ and a cross-section QS to a point S of $\theta(V_{\alpha,k})$. Then

$$f(x, y) = f(C_\alpha) + \delta_\alpha f_\alpha(x, y) \text{ in } \lambda(V_\alpha),$$

$$f(x, y) = f(C_{\alpha,k}) + \delta_{\alpha,k} f_{\alpha,k}(x, y) \text{ in } \lambda(V_{\alpha,k}).$$

These equations and $\delta_{\alpha,k} = \epsilon_{\alpha,k}\delta_\alpha$ give

$$f(x, y) = f(C_\alpha) + \delta_\alpha f_\alpha(C_{\alpha,k}) + \delta_\alpha \epsilon_{\alpha,k} f_{\alpha,k}(x, y)$$

in $\lambda(V_{\alpha,k})$. Now if $D^\alpha_{t_{\alpha,k}} \subset B(C_{\alpha,k})$, then $f_\alpha(x, y)$ increases monotonely from P to Q, by Theorem 41 and its corollary. Further $f_{\alpha,k}(x, y)$ increases from Q to S. Since also $\epsilon_{\alpha,k} = 1$, this implies that $f(x, y)$ varies monotonely along PQS. If $D^\alpha_{t_{\alpha,k}} \not\subset B(C_{\alpha,k})$, then $f_\alpha(x, y)$ decreases from P to Q and, since $\epsilon_{\alpha,k} = -1$, $\epsilon_{\alpha,k} f_{\alpha,k}(x, y)$ decreases from Q to S. Again $f(x, y)$ varies monotonely along PQS. The same holds for Q on a curve $C^*_{\alpha,k}$.

Finally take Q on C_1 and choose cross-sections QP and QS with P in $\theta(V_1)$, S in $\theta(V_1^*)$. Then $f(x, y) = f_1(x, y)$ increases from Q to P and $f(x, y) = -f_1^*(x, y)$ decreases from Q to S. Again $f(x, y)$ varies monotonely on PQS. $f(x, y)$ thus satisfies condition (3).

The locus $f(x, y) = c$ certainly consists of a set of curves of F. If there were more than countably many curves in the locus, then in some $\lambda(V_\alpha)$ or $\lambda(V_\alpha^*)$ there would be more than countably many. But that is impossible. For V_α is at most countable and, since by its definition $f_\alpha(x, y)$ takes on the value c at most once in $\theta(V_\alpha)$, the same holds for $f(x, y)$. There is a similar contradiction for the V_α^*. (2) thus holds and the theorem is established.

Remark 1. This theorem is a generalization of a theorem of E. Kamke (Mathematische Annalen, vol. 99(1928), p. 613).

Remark 2. If F is a regular family filling any open region R and if a function $f(x, y)$ exists with the properties (1), (2), (3) in R, then F is necessarily orientable. For take a point P on a curve C of F and an r-neighborhood $U(P)$. Map $\overline{U(P)}$ o-homeomorphically on an r-rectangle $|x'| \leqq 1$, $|y'| \leqq 1$, so that P has image on $y' = 0$. $f(x, y)$ then becomes a function f' of y' alone and is moreover monotone. We direct the line $y' = 0$ in such a way that, as we move

along $y' = 0$ in the positive direction, f' increases to the left. We orient the curve C at P accordingly. This gives a unique direction throughout P and a proper orientation of the whole family. We have thus:

COROLLARY TO THEOREM 42. *Every regular family F filling the plane is orientable.*

F can thus be parametrized by Whitney's function $f(p, t)$ (see §3.4 above).

BIBLIOGRAPHY

1. I. BENDIXSON, *Sur les courbes définies par des équations différentielles*, Acta Mathematica, vol. 24(1901), pp. 1–88.
2. G. D. BIRKHOFF, *Dynamical Systems*, American Mathematical Society Colloquium Publications, vol. 9, New York, 1927, especially Chapter VII, pp. 189–208.
3. L. E. J. BROUWER, *On continuous vector distributions on surfaces*, Akad. Wetensch. Amsterdam, Proceedings, vol. 11(1908), pp. 850–858; vol. 12(1909), pp. 716–734; vol. 13(1910), pp. 171–186.
4. A. DENJOY, *Sur les courbes définies par des équations différentielles à la surface du tore*, Journal de Mathématiques pures et appliquées, (9), vol. 11(1932), pp. 333–375.
5. E. GEORGE, *Topologie regulärer Kurvenscharen*, Deutsche Mathematik, vol. 4(1939), pp. 462–476. (See also Mathematical Reviews, vol. 1(1940), p. 29.)
6. B. v. KERÉKJÁRTÓ, *Vorlesungen über Topologie*, I, Berlin, 1923, §7, pp. 238–261.
7. H. KNESER, *Reguläre Kurvenscharen auf den Ringflächen*, Mathematische Annalen, vol. 91(1924), pp. 135–154.
8. H. POINCARÉ, *Les courbes définies par des équations différentielles*, Journal de Mathématiques pures et appliquées, (3), vol. 7(1881), pp. 375–422; (3), vol. 8(1882), pp. 251–296; (4), vol. 1(1885), pp. 167–244; (4), vol. 2(1886), pp. 151–217.
9. H. WHITNEY, *Regular families of curves*, Annals of Mathematics, (2), vol. 34(1933), pp. 244–270.

UNIVERSITY OF MICHIGAN.

THE FUNDAMENTAL SOLUTION OF THE PARABOLIC EQUATION

By F. G. Dressel

1. **Introduction.** The function

$$\frac{1}{[4\pi(y-\eta)]^{\frac{1}{2}n}} \exp\left(\frac{-\sum_{i=1}^{n}(x_i-\xi_i)^2}{4(y-\eta)}\right) \qquad (y > \eta) \tag{1}$$

is known as the fundamental solution of the parabolic equation

$$\Delta u - \frac{\partial u}{\partial y} = 0 \qquad \left(\Delta u = \sum_{i=1}^{n} \frac{\partial^2 u}{\partial x_i^2}\right).$$

Following a method of successive approximations introduced by Hadamard[1] for the case $n = 1$, Gevrey,[2] using the function (1) as the first approximation, showed the existence of a fundamental solution of the equation

$$\Delta u + \sum_{i=1}^{n} a_i \frac{\partial u}{\partial x_i} + au - \frac{\partial u}{\partial y} = 0. \tag{2}$$

If in equation (2) we replace Δu by an elliptic operator

$$H(u) = \sum_{i,k=1}^{n} \frac{\partial}{\partial x_i}\left(a_{ik}\frac{\partial u}{\partial x_k}\right),$$

then the function in (1) is no longer available as the first approximation of the fundamental solution of this new equation. For $n < 3$, this new equation can be transformed into the equation (2), but for $n > 2$, this is not the case. Thus for $n > 2$, the existence of a fundamental solution is not shown by Gevrey's method.

In case the a_{ij} in $H(u)$ are not functions of the variable y, Rothe[3] has shown that the equation

$$H(u) - \frac{\partial u}{\partial y} = 0$$

Received February 23, 1940; presented to the American Mathematical Society, February 24, 1940. The author is indebted to Mr. Willy Feller for suggesting this paper and for many valuable criticisms during its preparation.

[1] J. Hadamard, *Sur la solution fondamentale des équations aux dérivées partielles du type parabolique*, Paris Comptes Rendus, vol. 152(1911), pp. 1148–1149. For another treatment of this case see W. Feller, *Zur Theorie der stochastischen Prozesse*, Math. Annalen, vol. 113(1936–37), pp. 113–160.

[2] M. Gevrey, *Sur les équations aux dérivées partielles du type parabolique*, Journal de Mathématiques, (6), vol. 10(1913), pp. 105–148.

[3] E. Rothe, *Über die Grundlösung bei parabolischen Gleichungen*, Math. Zeitschrift, vol. 33(1931), pp. 488–504.

has a fundamental solution. He obtained his results by making use of the Green's functions of $H(u)$ and $H(u) - \lambda u$, so his methods are quite different from those of Gevrey.

In the present paper *we are concerned with the fundamental solution of the parabolic equation*

$$L(u) \equiv \sum_{i,j=1}^{n} a_{ij} \frac{\partial^2 u}{\partial x_i \partial x_j} + \sum_{i=1}^{n} a_i \frac{\partial u}{\partial x_i} + au - \frac{\partial u}{\partial y} = 0 \tag{3}$$

under the following assumptions: Let R be a bounded open region in the $x = (x_1, \cdots, x_n)$-space with the boundary R'; then for $\eta \leqq y \leqq y'$ and x in $R + R'$ the functions

$$\frac{\partial}{\partial y} a_{ij}; \qquad \frac{\partial^2}{\partial x_i \partial x_j} a_{ks}; \qquad a_i; \qquad a \qquad (a_{ij} = a_{ji}; i, j, k, s = 1, \cdots, n) \tag{4}$$

satisfy a Lipschitz condition[4] of order γ where $0 < \gamma \leqq 1$; also for these values of x, y

$$\sum_{i,j=1}^{n} a_{ij} \xi_i \xi_j$$

is a positive definite form. Under these conditions *we construct the fundamental solution of* (3) *in R, that is, a function $\Gamma(x, y; \xi, \eta)$ with the following properties*:

(j) For $y > \eta$ and for each pair of points x, ξ lying in R, the function $\Gamma(x, y; \xi, \eta)$ is a regular solution of equation (3).

(jj) If T is a subregion of R and $\varphi(x)$ is a function continuous in $R + R'$, then

$$\lim_{y \to \eta} \int_T \varphi(\xi) \Gamma(x, y; \xi, \eta) \, d\xi = \begin{cases} \varphi(x) & \text{if } x \text{ is a point interior to } T, \\ 0 & \text{if } x \text{ is a point interior to } R - T. \end{cases}$$

The methods we use to prove the existence of the fundamental solution are similar to those of Levi[5] for the elliptic equation. That is, we start with the principal part of Γ, say $Z(x, y; \xi, \eta)$, which is similar to (1), and write

$$\Gamma = Z + \int_\eta^y \int_R Z(x, y; s, \tau) f(s, \tau; \xi, \eta) \, ds \, d\tau.$$

Now put Γ in $L(u) = 0$, and we are left with an integral equation to determine $f(x, y; \xi, \eta)$.

2. Definitions and notation. In the determinant $| a_{ij} |$, let the co-factor of a_{ij} divided by $| a_{ij} |$ be denoted by A_{ij}; then the function

$$\sigma(x, y; \xi) = \sum_{i,j=1}^{n} A_{ij} \xi_i \xi_j$$

[4] See formula (ii).

[5] E. E. Levi, *Sulle equazioni lineari totalmente ellittiche alle derivate parziali*, Rend. del. Circ. Mat. Palermo, vol. 24(1907), pp. 275–317.

is a positive definite form for $\eta \leqq y \leqq y'$, x in $R + R'$. Therefore there exist two positive constants g and G such that

$$\text{(i)} \qquad gr^2 \leqq \sigma(x, y; x - \xi) \leqq Gr^2 \qquad \left(r^2 = \sum_{i=1}^{n} (x_i - \xi_i)^2\right).$$

We now define a function closely related to the function (1)

$$U(x, y; \xi, s) = \begin{cases} s^{-\frac{1}{2}n} \exp\left(-\dfrac{\sigma(x, y; \xi)}{4s}\right), & s > 0, \\ 0, & s \leqq 0. \end{cases}$$

The function $U(x, y, x - \xi, y - \eta)$ will also be denoted by

$$U(x, y/\xi, \eta).$$

We shall say *a function* $f(\xi, \tau)$ *belongs to the class* A *at the point* (x, y) if it satisfies the following three conditions:

$$\text{(1)} \qquad \int_{\eta}^{y'} \int_{R} |f(\xi, \tau)| \, d\xi \, d\tau < \infty,$$

where $\eta \leqq \tau \leqq y'$ and $d\xi = d\xi_1 \cdots d\xi_n$ is the volume element in R.

(2) $f(\xi, \tau)$ is bounded on any closed set for which $\tau > \eta$.

(3) $f(\xi, \eta)$ satisfies a Lipschitz condition of order γ $(0 < \gamma \leqq 1)$ at (x, y). That is, there exist constants $N, \delta, b > 0$ such that

$$\text{(ii)} \qquad |f(x, y) - f(s, \tau)| \leqq N[|y - \tau|^{\gamma} + \sum_{i=1}^{n} |x_i - s_i|^{\gamma}]$$

for $\eta < y - \delta \leqq \tau \leqq y + \delta$, $x - b \leqq s \leqq x + b$.

A *function* $H(x, y; \xi, \eta)$ *is said to be of order* $B(\alpha)$, if for $\eta \leqq y \leqq y'$, x in R there exist positive constants M and h such that

$$|H(x, y; \xi, \eta)| \leqq \frac{M}{(y - \eta)^{\alpha}} \exp(h\psi(x - \xi, y - \eta)),$$

where

$$\text{(5)} \qquad \psi(x, y) = -\frac{\sum_{i=1}^{n} x_i^2}{4y}.$$

Finally we note that if p, h are two positive constants with $0 < h < 1$, and α is a variable $\geqq 0$, then there exists a constant K such that

$$\text{(iii)} \qquad \alpha^{p} \exp(-\alpha) < K \exp(-h\alpha).$$

Any sum or product of a finite number of K's arising from property (iii) *of* $\exp(-\alpha)$ *shall be indicated by* (K).

3. **Differentiation of $V(x, y)$.** In this section we shall show that the function

$$V(x, y) = \int_\eta^y \int_R U(x, y/\xi, \tau) f(\xi, \tau)\, d\xi\, d\tau \tag{6}$$

has a derivative with respect to y ($y > \eta$), and second derivatives with respect to the x's, at each point (x, y) where f belongs to the class A.

Our first theorem is concerned with the differentiation of $V(x, y)$ with respect to y.

THEOREM 1. *If $f(\xi, \tau)$ belongs to class* A *at (x, y), and if $y > \eta$ and x is in R, then*

$$\begin{aligned}\frac{\partial}{\partial y} \int_\eta^y \int_R U(x, y/\xi, \tau) f(\xi, \tau)\, d\xi\, d\tau &= f(x, y) F(x, y) \\ &\quad + \int_\eta^y \int_R [f(\xi, \tau) - f(x, y)] \frac{\partial}{\partial y} U(x, y/\xi, \tau)\, d\xi\, d\tau \\ &\quad + f(x, y) \int_\eta^y \int_R \frac{\partial}{\partial y} U(x, y/\xi, \tau)\, d\xi\, d\tau,\end{aligned}$$

where $F(x, y)$ is defined in (8).

The integral involving the difference $[f(\xi, \eta) - f(x, y)]$ is to be interpreted as a multiple integral, while the last integral appearing in the theorem is to be evaluated as an iterated integral in the following sense:

$$\int_\eta^y \int_R \frac{\partial}{\partial y} U(x, y/\xi, \tau)\, d\xi\, d\tau = \lim_{\epsilon \to 0} \int_R \left\{ \int_\eta^{y-\epsilon} \frac{\partial}{\partial y} U(x, y/\xi, \tau)\, d\tau \right\} d\xi.$$

Proof. Write the difference ratio in the form

$$\begin{aligned}\frac{V(x, y + \Delta y) - V(x, y)}{\Delta y} &= \frac{1}{\Delta y} \int_y^{y+\Delta y} \int_R U(x, y + \Delta y/\xi, \tau) f(\xi, \tau)\, d\xi\, d\tau \\ &\quad + \frac{1}{\Delta y} \int_\eta^y \int_R [U(x, y + \Delta y/\xi, \tau) - U(x, y/\xi, \tau)] f\, d\xi\, d\tau \\ &= I + J.\end{aligned}$$

We consider the case $\Delta y > 0$. Let S be a sphere with center at x and of radius ρ. Assume ρ and δ taken so small that S lies in R and the Lipschitz condition on $f(\xi, \tau)$ holds for

$$y - \delta \leqq \tau \leqq y + \delta, \qquad \xi \subset S.$$

Denote the contribution to the integral I from the region $R - S$ by I_1, and the rest of I by I_2, then

$$|I_1| \leqq \frac{1}{\Delta y} \int_y^{y+\Delta y} \int_{R-S} (y + \Delta y - \tau)^{-\frac{1}{2}n} \exp\left(-\frac{g\rho^2}{4(y + \Delta y - \tau)}\right) |f|\, d\xi\, d\tau,$$

where $g\rho^2$ comes from property (i) of $\sigma(x, y; \xi)$. Using the property (iii) of $\exp(-\alpha)$ and the fact that f is bounded for $y > \eta$, we see that

$$\lim_{\Delta y \to 0} I_1 = 0. \tag{7}$$

In considering the other part of I, we write

$$\begin{aligned} I_2 = \frac{f(x, y)}{\Delta y} \int_y^{y+\Delta y} \int_S U(x, y + \Delta y/\xi, \tau)\, d\xi\, d\tau \\ + \frac{1}{\Delta y} \int_y^{y+\Delta y} \int_S U(x, y + \Delta y/\xi, \tau)[f(\xi, \tau) - f(x, y)]\, d\xi\, d\tau \\ = f(x, y) I_2^1 + I_2^2. \end{aligned}$$

If the

$$\lim_{\Delta y \to 0} I_2^1$$

exists, then, since f is continuous at (x, y), the limit of I_2^2 can be made arbitrarily small with δ and ρ. It then follows from (7) that I and $f(x, y)I_2^1$ have the same limit.

In the integral I_2^1 introduce polar coördinates defined by

$$\begin{aligned} \xi_1 - x_1 &= r \cos \theta_1, \\ \xi_2 - x_2 &= r \sin \theta_1 \cos \theta_2, \\ &\cdots\cdots\cdots\cdots \\ \xi_{n-1} - x_{n-1} &= r \sin \theta_1 \cdots \sin \theta_{n-2} \cos \theta_{n-1}, \\ \xi_n - x_n &= r \sin \theta_1 \cdots \sin \theta_{n-1}, \\ r^2 &= \sum_{i=1}^{n} (x_i - \xi_i)^2. \end{aligned}$$

When this transformation is applied to the function $\sigma(x, y; x - \xi)$, r^2 can be factored out. Denote the remaining factor by $\varphi(x, y; \theta)$, then

$$\sigma(x, y; x - \xi) = r^2 \varphi(x, y; \theta).$$

Following the change to polar coördinates in I_2^1, write

$$\tfrac{1}{2} r [\varphi(x, y; \theta)]^{\frac{1}{2}} [y + \Delta y - \tau]^{-\frac{1}{2}} = w,$$

then

$$I_2^1 = \frac{1}{\Delta y} \int_y^{y+\Delta y} \int_0^{2\pi} \int_0^{\pi} \cdots \int_0^{\pi} \int_0^{\beta} 2^n \varphi^{-\frac{1}{2}n} H(\theta) w^{n-1} \exp(-w^2)\, dw\, d\theta\, d\tau,$$

where

$$H(\theta) = \sin^{n-2}\theta_1 \sin^{n-3}\theta_2 \cdots \sin\theta_{n-2} \qquad (0 \leqq \theta_{n-1} \leqq 2\pi,\ 0 \leqq \theta_k \leqq \pi;\ k = 1, \cdots, n-2),$$

$$\beta = \tfrac{1}{2} \rho \varphi^{\frac{1}{2}} (y + \Delta y - \tau)^{-\frac{1}{2}}.$$

From the above representation of I_2^1 the limit is seen to exist, and we have

$$\lim_{\Delta y \to 0} I = \lim_{\Delta y \to 0} f(x, y) I_2^1 = F(x, y) f(x, y).$$

The function $F(x, y)$ is defined by

$$F(x, y) = 2^n \int_0^{2\pi} \int_0^{\pi} \cdots \int_0^{\pi} [\varphi(x, y; \theta)]^{-\frac{1}{2}n} H(\theta)\, d\theta_1 \cdots d\theta_{n-1} \tag{8}$$
$$\cdot \int_0^{\infty} w^{n-1} \exp(-w^2)\, dw.$$

Returning now to J, write it as three integrals

$$\begin{aligned} J &= \int_{\eta}^{y-\delta} \int_R + \int_{y-\delta}^{y} \int_{R-S} \\ &\qquad + \int_{y-\delta}^{y} \int_S \frac{1}{\Delta y} [U(x, y + \Delta y/\xi, \tau) - U(x, y/\xi, \tau)] f(\xi, \tau)\, d\xi\, d\tau \\ &= J_1 + J_2 + J_3. \end{aligned}$$

In the first integral $y - \tau \geqq \delta > 0$, and in the second integral $\sigma \geqq g\rho^2 > 0$, hence it readily follows that

$$\lim_{\Delta y \to 0} (J_1 + J_2) = \int_{\eta}^{y-\delta} \int + \int_{y-\delta}^{y} \int_{R-S} \frac{\partial}{\partial y} U(x, y/\xi, \tau) f(\xi, \tau)\, d\xi\, d\tau.$$

Write J_3 as two integrals $J_3^1 + J_3^2$. The $f(\xi, \tau)$ in J_3 is replaced by the difference $[f(\xi, \tau) - f(x, y)]$ to obtain J_3^1, and in J_3^2 we have $f(x, y)$ in place of $f(\xi, \tau)$. We first show that J_3^1 can be made arbitrarily small with δ. Using the theorem of the mean, we write J_3^1 in the form

$$J_3^1 = \int_{y-\delta}^{y} \int_S U_y(x, y + h\Delta y/\xi, \tau)[f(\xi, \tau) - f(x, y)]\, d\xi\, d\tau \qquad (0 < h < 1), \tag{9}$$

where

$$U_y = \frac{\partial}{\partial y} U.$$

Making use of the Lipschitz condition of order γ on f, we see that the integrand of J_3^1 is a function of order $B(\frac{1}{2}(n - \gamma) + 1)$. Using this fact in J_3^1, we then change to polar coördinates. Property (iii) is now used, and we see there is a constant M such that

$$| J_3^1 | \leqq M \int_{y-\delta}^{y} \int_S \frac{1}{r^{\beta}(y + h\Delta y - \tau)^s}\, dr\, d\theta\, d\tau, \qquad s = \frac{3 - \gamma - \beta}{2}.$$

Dropping $h\Delta y$ and taking $\beta = 1 - \frac{1}{2}\gamma$, we see that the above integral exists; hence J_3^1 approaches zero with δ.[6]

[6] For $\Delta y < 0$, the integral corresponding to J_3^1 has to be handled in a slightly different manner.

We remark that if h is put equal to zero in the integral in (9), the same method that we used on J_3^1 shows that this integral approaches zero with δ. Therefore, we may pass to the limit under the integral sign in J_3^1.

Turning to J_3^2, we write it as two integrals. In the integral that contains $U(x, y + \Delta y/\xi, \tau)$, change the variable of integration τ to $\tau + \Delta y$; then we can write J_3^2 in the form

$$J_3^2 = \frac{f(x,y)}{\Delta y} \int_{y-\delta-\Delta y}^{y-\delta} \int_S U(x, y + \Delta y; x - \xi, y - \tau)\, d\xi\, d\tau$$

$$+ \frac{f(x,y)}{\Delta y} \int_{y-\delta}^{y-\Delta y} \int_S \{U(x, y + \Delta y; x - \xi, y - \tau) - U(x, y; x - \xi, y - \tau)\}\, d\xi\, d\tau$$

$$- \frac{f(x,y)}{\Delta y} \int_{y-\Delta y}^{y} \int_S U(x, y/\xi, \tau)\, d\xi\, d\tau = S_1 + S_2 + S_3.$$

The limit of S_1 is easily obtained since the integrand is continuous. In S_3 change the variable τ to $\tau - \Delta y$, and then its limit can be found in the same manner as the limit of I_2^1. Applying the theorem of the mean to S_2, we have

$$S_2 = f(x,y) \int_{y-\delta}^{y-\Delta y} \int_S \frac{-1}{4(y-\tau)}\, \sigma_y(x, y + \theta\Delta y; x - \xi)\, U(x, y + \theta\Delta y; x - \xi, y - \tau)\, d\xi\, d\tau$$
$$(0 < \theta < 1).$$

It is no trouble to show that the limit may be taken under the sign of the integral, since Δy enters only in the continuous functions

$$A_{ij}, \qquad \frac{\partial}{\partial y} A_{ij}.$$

The limit of J_3^2 is thus seen to be

$$\lim_{\Delta y \to 0} J_3^2 = f(x,y) \int_S U(x, y; x - \xi, \delta)\, d\xi$$

$$- f(x,y) \int_{y-\delta}^{y} \int_S \frac{1}{4(y-\tau)}\, \sigma_y(x, y; x - \xi)\, U(x, y/\xi, \tau)\, d\xi\, d\tau - f(x,y)F(x,y).$$

This completes the proof of Theorem 1 since the limit for J_3^2 is the same as the following one:

$$f(x,y) \int_{y-\delta}^{y} \int_S U_y(x, y/\xi, \tau)\, d\xi\, d\tau = f(x,y) \int_S U(x, y, x - \xi, \delta)\, d\xi$$

$$- f(x,y) \int_{y-\delta}^{y} \int_S \frac{1}{4(y-\tau)}\, \sigma_y(x, y, x - \xi)\, U(x, y/\xi, \tau)\, d\xi\, d\tau$$

$$- \lim_{\epsilon \to 0} f(x,y) \int_S U(x, y, x - \xi, \epsilon)\, d\xi.$$

Some properties of the function $F(x, y)$ that we shall need later are now stated in the following lemma.

LEMMA 1. *The function $F(x, y)$ defined in* (8) *is differentiable with respect to each of its arguments, and satisfies the inequalities*

$$(4\pi/G)^{\frac{1}{2}n} \leqq F(x, y) \leqq (4\pi/g)^{\frac{1}{2}n}, \quad \eta \leqq y \leqq y', \quad x \text{ in } R + R',$$

where g and G come from property (i) *of* σ.

This lemma follows from the definition of $F(x, y)$ and the properties of σ.

Our second theorem is concerned with the differentiation with respect to the x's of the function $V(x, y)$ defined in (6).

THEOREM 2. *If $f(\xi, \tau)$ belongs to class* A *at* (x, y), *then*

$$\frac{\partial V}{\partial x_i} = \int_\eta^y \int_R \frac{\partial}{\partial x_i} U(x, y/\xi, \tau) f(\xi, \tau)\, d\xi\, d\tau, \tag{10}$$

and

$$\frac{\partial^2 V}{\partial x_i \partial x_j} = \int_\eta^y \int_R [f(\xi, \tau) - f(x, y)] \frac{\partial^2}{\partial x_i \partial x_j} U(x, y/\xi, \tau)\, d\xi\, d\tau + f(x, y) \int_\eta^y d\tau \int_R \frac{\partial^2 U}{\partial x_i \partial x_j} d\xi. \tag{11}$$

All integrals appearing in Theorem 2 are to be considered as multiple integrals except the last one, and it is to be evaluated as an iterated integral in the following order:

$$\int_\eta^y d\tau \int_R \frac{\partial^2}{\partial x_i \partial x_j} U\, d\xi = \lim_{\epsilon \to 0} \int_\eta^{y-\epsilon} \left\{ \int_R \frac{\partial^2}{\partial x_i \partial x_j} U\, d\xi \right\} d\tau.$$

Before taking up the proof of Theorem 2, we shall prove a useful lemma.

Let x be a point in R and P be the region defined by

$$x_i - b \leqq \xi_i \leqq x_i + b \qquad (i = 1, \cdots, n).$$

In what follows, we assume that b and another positive constant δ have been taken so small that $P \subset R$ and the Lipschitz condition on $f(\xi, \tau)$ at (x, y) holds throughout the space

$$P, \eta < y - \delta \leqq \tau \leqq y.$$

LEMMA 2. *Let $H(x - \xi) = (x_1 - \xi_1)^{\alpha_1} \cdots (x_n - \xi_n)^{\alpha_n}$, where the α_i are positive integers or zero and $\alpha_1 + \cdots + \alpha_n = \alpha$; also let $f(\xi, \tau)$ belong to class* A *at (x, y). Then the function*

$$S(x, y) = \int_{y-\delta}^y \int_P \frac{H(x - \xi)}{(y - \tau)^\beta} U(x, y/\xi, \tau) f(\xi, \tau)\, d\xi\, d\tau$$

has a derivative with respect to x_i at (x, y) equal to

$$\frac{\partial S}{\partial x_i} = \int_{y-\delta}^y \int_P \frac{f(\xi, \tau) - f(x, y)}{(y - \tau)^\beta} \frac{\partial}{\partial x_i} [HU]\, d\xi\, d\tau + f(x, y) \int_{y-\delta}^y d\tau \int_P \frac{\partial}{\partial x_i} [HU]\, d\xi,$$

if $2\beta < \alpha + \gamma + 1$. Here γ is the order of the Lipschitz condition on $f(\xi, \tau)$.

Consider the case $i = 1$ with $\Delta x_i = \Delta x > 0$. Let $x' = (x_1 + \Delta x, x_2, \cdots, x_n)$, and write

$$\frac{S(x', y) - S(x, y)}{\Delta x} = \frac{1}{\Delta x} \int_{y-\delta}^{y} \int_P \frac{H(x' - \xi)}{(y - \tau)^\beta} [U(x', y; x' - \xi, y - \tau)$$

$$- U(x, y; x' - \xi, y - \tau)] f(\xi, \tau)\, d\xi\, d\tau$$

$$+ \frac{1}{\Delta x} \int_{y-\delta}^{y} \int_P \frac{1}{(y - \tau)^\beta} [H(x' - \xi) U(x, y; x' - \xi, y - \tau)$$

$$- H(x - \xi) U(x, y/\xi, \tau)] f(\xi, \tau)\, d\xi\, d\tau = J(\delta) + I(\delta).$$

After applying the theorem of the mean in $J(\delta)$, we see that the integrand is a function of order $B(s)$, where $2s = (2\beta + n - \alpha)$; hence

$$(12) \qquad | J(\epsilon) | \leqq M \int_{y-\epsilon}^{y} \int_P \frac{1}{(y - \tau)^s} \exp [h\psi(x' - \xi, y - \tau)]\, d\xi\, d\tau, \qquad 0 < h < g,$$

where the function ψ is defined in (5). Since $0 < \gamma \leqq 1$, conditions of the lemma insure us that $2\beta < \alpha + 2$. From this it follows that the above integral exists, and therefore $J(\epsilon)$ approaches zero with ϵ. From this we conclude that we may pass to the limit under the integral sign in $J(\delta)$.

Write $I(\delta) = I_1(\delta) + I_2(\delta)$. $I_1(\delta)$ is obtained from $I(\delta)$ by replacing $f(\xi, \tau)$ by the difference $[f(\xi, \tau) - f(x, y)]$, while $I_2(\delta)$ is obtained by replacing $f(\xi, \tau)$ by $f(x, y)$. We first show $I_1(\epsilon)$ is small with ϵ.

In $I_1(\epsilon)$ break the integration with respect ξ_1 into the three parts

$$x_1 - b \leqq \xi_1 \leqq x_1, \quad x_1 \leqq \xi_1 \leqq x_1 + 2\Delta x, \quad x_1 + 2\Delta x \leqq \xi_1 \leqq x_1 + b,$$

and call the contribution to $I_1(\epsilon)$ of these $I_1'(\epsilon)$, $I_1''(\epsilon)$, $I_1'''(\epsilon)$ respectively.

In $I_1'(\epsilon)$ apply the theorem of the mean. Making use of the Lipschitz condition on f and also the fact that over the interval $x_1 - b \leqq \xi_1 \leqq x_1$ one has

$$0 \leqq x_1 - \xi_1 \leqq x_1 + \theta \Delta x - \xi_1 \qquad (0 < \theta < 1),$$

then $| I_1'(\epsilon) |$ is seen to be dominated by the same type of integral as appears in (12). The dominating integral here has x instead of x' and $2s$ has the value $(2\beta + n + 1 - \alpha - \gamma)$. However, the conditions in the lemma still insure $I_1'(\epsilon)$ is $o(1)$ with ϵ. $I_1'''(\epsilon)$ can be treated like $I_1'(\epsilon)$ since

$$0 \leqq \xi_1 - x_1 \leqq 2(\xi_1 - x_1 - \theta \Delta x) \qquad (0 < \theta < 1),$$

over the interval $x_1 + 2\Delta x \leqq \xi_1 \leqq x_1 + b$.

In $I_1''(\epsilon)$ we do not use the theorem of the mean, so we may write

$$| I_1''(\epsilon) | \leqq \frac{(K)N}{\Delta x} \int_{y-\epsilon}^{y} \int_{P'} \frac{1}{(y - \tau)^s} [\exp (h\psi(x' - \xi, y - \tau))$$

$$+ \exp (h\psi(x - \xi, y - \tau))]\, d\xi\, d\tau$$

$$+ \frac{(K)N}{\Delta x} \int_{y-\epsilon}^{y} \int_{P'} \frac{| x_1 - \xi_1 |^\gamma}{(y - \tau)^v} \exp (h\psi(x' - \xi, y - \tau))\, d\xi\, d\tau \qquad (0 < h < g),$$

where

$$2s = (2\beta + n - \alpha - \gamma), \qquad 2v = (2\beta + n - \alpha),$$

N comes from the Lipschitz condition on f, and P' is the subset of P where $x_1 \leqq \xi_1 \leqq x_1 + 2\Delta x$. We increase the integrand of the first integral on the right by replacing $(x_1 + \Delta x - \xi_1)$ and $(x_1 - \xi_1)$ by zero. The range of ξ_1 then takes care of Δx, and the resulting integral is $o(1)$ with ϵ. In the second integral multiply numerator and denominator of its integrand by

$$| x_1 + \Delta x - \xi_1 |^{1-\gamma},$$

then

$$| x_1 - \xi_1 |^{\gamma} | x_1 + \Delta x - \xi_1 |^{1-\gamma} < 2\Delta x$$

over P'. After canceling out the Δx's we see that the remaining integral is $o(1)$ with ϵ, since $\gamma > 0$.

This completes the proof that $I_1(\epsilon) = o(1)$ with ϵ. If in $I_1(\epsilon)$ we replace the incremental ratio by the derivative, it is easy to show that this function is $o(1)$ with ϵ. Thus in $I_1(\delta)$ we may pass to the limit under the sign of integration.

To complete the proof of the lemma, we must show that

$$\text{(13)} \qquad \lim_{\Delta x \to 0} I_2(\delta) = -f(x, y) \int_{y-\delta}^{y} d\tau \int_P \frac{\partial}{\partial \xi_1} [H(x - \xi) U(x, y/\xi, \tau)] \, d\xi.$$

Separate $I_2(\delta)$ into two integrals. In the integral in which x' is concerned, change the variable of integration ξ_1 to $\xi_1 + \Delta x$, then we have

$$I_2(\delta) = \frac{f(x, y)}{\Delta x} \left\{ \int_{y-\delta}^{y} \int_{P_1} \int_{x_1 - b - \Delta x}^{x_1 - b} - \int_{y-\delta}^{y} \int_{P_1} \int_{x_1 + b - \Delta x}^{x_1 + b} \right\} \frac{H(x - \xi)}{(y - \tau)^{\beta}} U(x, y/\xi, \tau) \, d\xi \, d\tau,$$

where P_1 is the set defined by $x_j - b \leqq \xi_j \leqq x_j + b$ $(j = 2, \cdots, n)$. For $\Delta x < \frac{1}{2}b$, the function σ in each of the above integrals is greater than the positive constant $\frac{1}{4}gb^2$, hence the integrand is continuous, and we have

$$\lim_{\Delta x \to 0} I_2(\delta) = f(x, y) \int_{y-\delta}^{y} \int_{P_1} \left[\frac{H(x'' - \xi)}{(y - \tau)^{\beta}} U(x, y; x'' - \xi, y - \tau) \right.$$
$$\left. - \frac{H(x' - \xi)}{(y - \tau)^{\beta}} U(x, y; x' - \xi, y - \tau) \right] d\xi' \, d\tau,$$

where

$$d\xi' = d\xi_2 \, d\xi_3 \cdots d\xi_n,$$

$$x'' - \xi = (b, x_2 - \xi_2, \cdots, x_n - \xi_n), \qquad x' - \xi = (-b, x_2 - \xi_2, \cdots, x_n - \xi_n).$$

Because of the way x and ξ appear in H and U, the above integral equals the integral in (13).

With the lemma proved, we now return to the consideration of Theorem 2. Write $V(x, y)$ defined in (6) in the form

$$\text{(14)} \qquad V(x, y) = \int_{\eta}^{y-\delta} \int_R + \int_{y-\delta}^{y} \int_{R-P} + \int_{y-\delta}^{y} \int_P U f \, d\xi \, d\tau.$$

In the first two integrals on the right in (14) either $y - \tau \geqq \delta > 0$ or $\sigma \geqq gb^2 > 0$; hence they can be differentiated under the integral sign at least twice with respect to the x's. The last integral in (14) is the type to which Lemma 2 applies. Therefore, we can state that Theorem 2 holds. It is to be noted that the integral appearing in (10) exists as a multiple integral, so we do not need to use the difference $[f(\xi, \tau) - f(x, y)]$.

4. **The order of the terms in** $L(u)$. The first derivative of $U(x, y/\xi, \eta)$ with respect to x_i is a function of order $B(\frac{1}{2}(n + 1))$, and the second derivative can be written in the form

$$(15)\quad \frac{\partial^2 U}{\partial x_i \partial x_k} = \left[-\frac{A_{ik}}{2(y-\eta)} + \frac{\sum_{j=1}^{n} \sum_{p=1}^{n} A_{ij} A_{kp} (x_j - \xi_j)(x_p - \xi_p)}{4(y - \eta)^2} \right] U + W_1(x, y; \xi, \eta),$$

where $W_1(x, y; \xi, \eta)$ is a function of order $B(\frac{1}{2}(n + 1))$. Similarly we have

$$(16)\quad \frac{\partial U}{\partial y} = \left[-\frac{n}{2(y - \eta)} + \frac{\sigma}{4(y - \eta)^2} \right] U + W_2(x, y; \xi, \eta).$$

The function W_2 is of order $B(\frac{1}{2}(n + 1))$.

Making use of the definition of the A_{ij}, we have from (15) that

$$(17)\quad \sum_{i,k=1}^{n} a_{ij} \frac{\partial^2 U}{\partial x_i \partial x_k} = \left[-\frac{n}{2(y - \eta)} + \frac{\sigma}{4(y - \eta)^2} \right] + \text{a function of order } B\left(\frac{n+1}{2}\right).$$

It is evident from (16) and (17) that, if in $L(u)$ defined in (3) we substitute U, the resulting function is of order $B(\frac{1}{2}(n + 1))$.

5. **The integral equation.** Since the function $F(x, y)$ defined in (8) is bounded away from zero in the space under consideration, we may define the function

$$Z(x, y; \xi, \eta) = \frac{U(x, y/\xi, \eta)}{F(\xi, \eta)}.$$

Using this function, define the function $\Gamma(x, y; \xi, \eta)$ as

$$\Gamma(x, y; \xi, \eta) = Z(x, y; \xi, \eta) + \int_{\eta}^{y} \int_{R} Z(x, y; s, \tau) f(s, \tau; \xi, \eta)\, ds\, d\tau,$$

where we assume for fixed (ξ, η) that the function $f(s, \tau; \xi, \eta)$ is a function of class A for each point (x, y) such that

$$\eta < y, \qquad x \subset R.$$

Substituting Γ in $L(u)$ and using Theorems 1 and 2, we have

$$L(\Gamma) = L(Z) - f(x, y; \xi, \eta) + \int_\eta^y \int_R L(U(x, y; s, \tau)) \left[\frac{f(s, \tau; \xi, \eta)}{F(s, \tau)} - \frac{f(x, y; \xi, \eta)}{F(x, y)}\right] ds\, d\tau + \frac{f(x, y; \xi, \eta)}{F(x, y)} \int_\eta^y \int_R L(U)\, ds\, d\tau.$$

In the preceding article we showed that $L(U)$ was a function of order $B(\frac{1}{2}(n + 1))$; hence we have

$$L(\Gamma) = L(Z) - f(x, y; \xi, \eta) + \int_\eta^y \int_R L(Z(x, y; s, \tau)) f(s, \tau; \xi, \eta)\, ds\, d\tau.$$

Setting $L(\Gamma) = 0$, we are left with the following integral equation to determine f:

$$f(x, y; \xi, \eta) = L(Z(x, y; \xi, \eta)) + \int_\eta^y \int_R L(Z(x, y; s, \tau)) f(s, \tau; \xi, \eta)\, ds\, d\tau, \tag{18}$$

where $L(Z)$ satisfies a relation of the following type

$$|L(Z(x, y; \xi, \eta))| \leqq \frac{M}{(y - \eta)^{\frac{1}{2}(n+1)}} \exp[h\psi(x - \xi, y - \eta)] \qquad (0 < h < g), \tag{19}$$

for points x, ξ in $R + R'$, and $\eta \leqq y \leqq y'$.

We shall show that the integral equation (18) has as a solution a function of class A at each point (x, y) such that

$$\eta < y \leqq y', \qquad x \subset R.$$

First, however, we shall need some lemmas with reference to the following type of function

$$W(x, y; \xi, \eta; \alpha, h) = \frac{1}{(y - \eta)^\alpha} \exp[h\psi(x - \xi, y - \eta)].$$

LEMMA 3. *If $0 \leqq \alpha, \beta < \frac{1}{2}n + 1$, and h, l are positive constants, then*

$$\int_\eta^y \int_R W(x, y; s, \tau; \alpha, h) W(s, \tau; \xi, \eta; \beta, l)\, ds\, d\tau \leqq K(\alpha, \beta, l, h) W(x, y; \xi, \eta; \beta + \alpha - \tfrac{1}{2}n - 1, p),$$

where $4p = \min(h, l)$ and $K(\alpha, \beta, l, h)$ is a constant depending on α, β, l, and h.

Let $m(x, \xi)$ be the distance function

$$m(x, \xi) = \left[\sum_{i=1}^n (x_i - \xi_i)^2\right]^{\frac{1}{2}},$$

and write for short

$$r = m(x, \xi), \qquad \rho = m(s, \xi), \qquad \delta = m(x, s);$$

then

$$(r - \rho)^2 \leqq \delta^2, \qquad (r - \delta)^2 \leqq \rho^2.$$

Break up the integral in Lemma 3 into the sum of two integrals, $I_1 + I_2$. In I_1 the τ integration is over $\eta \leqq \tau \leqq \frac{1}{2}(y + \eta)$, and in I_2, τ ranges over the remainder of the interval (η, y).

In I_1, after changing to polar coördinates[7] with center at ξ, and making use of $(r - \rho)^2 \leqq \delta^2$, we have

$$I_1 \leqq H \int_\eta^{\frac{1}{2}(y+\eta)} \int_0^\infty \frac{\rho^{n-1}}{(y-\tau)^\alpha(\tau-\eta)^\beta} \exp\left[-\frac{h(r-\rho)^2}{4(y-\tau)} - \frac{l\rho^2}{4(\tau-\eta)}\right] d\rho\, d\tau,$$

where

$$H = \int_0^{2\pi} \int_0^\pi \cdots \int_0^\pi H(\theta)\, d\theta. \tag{20}$$

Replace $(y - \tau)^\alpha$ by its minimum value over the range of integration, then use property (iii) to obtain

$$I_1 \leqq \frac{H(K)2^{\alpha+n-2}}{(y-\eta)^\alpha l^{\frac{1}{2}(n-2)}} \int_\eta^{\frac{1}{2}(y+\eta)} \int_0^\infty \frac{\rho}{(\tau-\eta)^{\beta-\frac{1}{2}(n-2)}} \exp\left[-\frac{h(r-\rho)^2}{4(y-\tau)} - \frac{l\rho^2}{8(\tau-\eta)}\right] d\rho\, d\tau.$$

Over the range of integration $0 \leqq \rho \leqq \frac{1}{2}r$ make use of the first of the following inequalities, over the range $\frac{1}{2}r \leqq \rho < \infty$ use the second inequality:

$$\exp\left[-\frac{h(r-\rho)^2}{4(y-\tau)}\right] \leqq \exp\left[-\frac{hr^2}{16(y-\eta)}\right], \qquad 0 \leqq \rho \leqq \tfrac{1}{2}r, \qquad \eta \leqq \tau \leqq \tfrac{1}{2}(y+\eta),$$
$$\leqq 1,$$

and one can readily obtain

$$I_1 \leqq H(K) \frac{2^{\alpha+\beta+1+\frac{1}{2}n}}{l^{\frac{1}{2}n}(n-2\beta+2)} W(x, y; \xi, \eta; \alpha + \beta - \tfrac{1}{2}n - 1, p).$$

We can obtain a similar inequality for I_2 by using polar coördinates with center at x, and the fact that $(r - \delta)^2 \leqq \rho^2$. Thus the lemma holds.

Lemma 4. *If* $0 \leqq \alpha < \frac{1}{2}n + 1$, $\beta > -1$, $l > 0$, *then*

$$\int_\eta^y \int_R W(x, y; s, \tau; \alpha, l)(\tau - \eta)^\beta\, ds\, d\tau$$
$$\leqq \frac{1}{2}\left(\frac{4}{l}\right)^{\frac{1}{2}n} H\Gamma(\tfrac{1}{2}n)B(\tfrac{1}{2}n - \alpha + 1, \beta + 1)(y - \eta)^{\beta-\alpha+\frac{1}{2}n+1},$$

where H *is defined in* (20) *and* Γ, B *are the gamma and beta functions.*

The details of the proof are omitted.

[7] We give the proof for $n > 1$.

6. **Solution of the integral equation.** Returning now to the integral equation in (18), we have by successive substitutions

$$
\begin{aligned}
f(x, y; \xi, \eta) &= L(Z(x, y; \xi, \tau)) \\
&\quad + \int_\eta^y \int_R L(Z(x, y; s, \tau))L(Z(s, \tau; \xi, \eta))\, ds\, d\tau + \cdots .
\end{aligned} \tag{21}
$$

Using inequality (19), we find that

$$
\begin{aligned}
|f(x, y; \xi, \eta)| &\leqq MW\left(x, y; \xi, \eta; \frac{n+1}{2}, h\right) \\
&\quad + M^2 \int_\eta^y \int_R W\left(x, y; s, \tau; \frac{n+1}{2}, h\right) W\left(s, \tau; \xi, \eta; \frac{n+1}{2}, h\right) ds\, d\tau + \cdots .
\end{aligned} \tag{22}
$$

Referring to the $K(\alpha, \beta, l, h)$ of Lemma 3, write

$$
k = K\left(\frac{n+1}{2}, \frac{n+1}{2}, h, h\right) K\left(\frac{n+1}{2}, \frac{n}{2}, h, \frac{h}{4}\right) \cdots K\left(\frac{n+1}{2}, 1, h, \frac{h}{4^{n-1}}\right),
$$

and let the m-th term of (22) be denoted by $I(m)$. Then making use of Lemma 3, we have for $m > n$

$$
\begin{aligned}
I(m+2) \leqq M^{m+2} k \int_\eta^y \int_R W\left(x, y; s^1, \tau_1; \frac{n+1}{2}, h\right) \int_\eta^{\tau_1} \int_R \cdots \int_\eta^{\tau_i} \int_R \\
\cdot W\left(s^i, \tau_i; s^{i+1}, \tau_{i+1}; \frac{n+1}{2}, h\right) (\tau_{i+1} - \eta)^{-\frac{1}{2}} (ds^{i+1}\, d\tau_{i+1}) \cdots (ds^1\, d\tau_1),
\end{aligned}
$$

where $i = m - n$. We may now use Lemma 4 on the above integral. Let us call

$$
L = \frac{1}{2}\left(\frac{4}{h}\right)^{\frac{1}{2}n} H\Gamma(\tfrac{1}{2}n)\Gamma(\tfrac{1}{2});
$$

then

$$
I(m+2) \leqq \frac{M^{m+2} k L^{m-n+1} \Gamma(\frac{1}{2})}{\Gamma(\frac{1}{2}(m-n)+1)} (y - \eta)^{\frac{1}{2}(m-n)}.
$$

From the above inequality and the ratio test the series in (22) is seen to be absolutely convergent. Hence it readily follows that the function defined by (21) satisfies the integral equation (18). In order that $\Gamma(x, y; \xi, \eta)$ satisfy $L(u) = 0$, we must show that $f(x, y; \xi, \eta)$ satisfies a Lipschitz condition of order $0 < \gamma \leqq 1$.

7. **The Lipschitz condition.** Again let

$$
H(x - \xi) = (x_1 - \xi_1)^{\alpha_1} \cdots (x_n - \xi_n)^{\alpha_n},
$$

the α_i being positive integers or zero, then the terms of $L(Z(x, y; \xi, \eta))$ are all of the form

$$\frac{B(x, y)H(x - \xi)}{(y - \eta)^{\beta}} Z(x, y; \xi, \eta), \tag{23}$$

where each $B(x, y)$ is a function of

$$a_{ij}, \qquad \frac{\partial}{\partial x_k} a_{ij}, \qquad \frac{\partial^2}{\partial x_k\, \partial x_m} a_{ij}, \qquad \frac{\partial}{\partial y} a_{ij}, \qquad a_i, \qquad a.$$

Because of the conditions on these functions and the way they enter $B(x, y)$, we see $B(x, y)$ satisfies a Lipschitz condition of order $\gamma > 0$ at each point of the space under consideration. The function $H(x - \xi)/(y - \eta)^{\beta}$ of (23), in terms of the symbol $[\alpha, \beta] = (\alpha_1 + \cdots + \alpha_n, \beta)$ is one of the following five possibilities

$$(0, 0), \qquad (1, 1), \qquad (2, 1), \qquad (3, 2), \qquad (4, 2). \tag{24}$$

For $y > \eta$ the function (23), with $B(x, y)$ replaced by 1, is differentiable with respect to each of its arguments. Thus to show $f(x, y; \xi, \eta)$ of (21) satisfies a Lipschitz condition, we need only consider the integrated terms. If γ_1 is the order of the Lipschitz condition satisfied by the $B(x, y)$, then the next two lemmas will assure us that $f(x, y; \xi, \eta)$ satisfies a Lipschitz condition of order $\gamma = \min(\gamma_1, \frac{1}{4})$.

LEMMA 5. *Let*

$$I(x, y) = \int_{\eta}^{y} \int_{R} \frac{H(x - s)}{(y - \tau)^{\beta}} U(x, y/s, \tau) Y(s, \tau)\, ds\, d\tau;$$

then with $z \geqq y > \eta$ there is a constant K such that

$$\begin{aligned} &| I(x, z) - I(x, y) | \\ &\leqq | z - y |^{\frac{1}{4}} \frac{K}{(y - \eta)^{\frac{1}{2}n} + 1} \left[\int_{\eta}^{y'} \int_{R} | Y(s, \tau) |\, ds\, d\tau + \max_{2\tau \geqq y + \eta} | Y(s, \tau) | \right], \end{aligned} \tag{25}$$

if $[\alpha, \beta]$ has any of the values in the set (24). *In taking the* $\max |Y|$ *the s are restricted to $R + R'$.*

Proof. Write the difference in (25) in the following form

$$\begin{aligned} I(x, z) - I(x, y) &= \int_{y}^{z} \int_{R} \frac{H(x - s)}{(z - \tau)^{\beta}} U(x, z/s, \tau) Y(s, \tau)\, ds\, d\tau \\ &+ \int_{\eta}^{y} \int_{R} H(x - s) \left[\frac{1}{(z - \tau)^{\frac{1}{2}n + \beta}} - \frac{1}{(y - \tau)^{\frac{1}{2}n + \beta}} \right] \exp\left(- \frac{\sigma(x, y; x - s)}{4(y - \tau)} \right) Y\, ds\, d\tau \\ &+ \int_{\eta}^{y} \int_{R} \frac{H(x - s)}{(z - \tau)^{\frac{1}{2}n + \beta}} \left[\exp\left(- \frac{\sigma(x, z; x - s)}{4(z - \tau)} \right) - \exp\left(- \frac{\sigma(x, y; x - s)}{4(z - \tau)} \right) \right] Y\, ds\, d\tau \\ &+ \int_{\eta}^{y} \int_{R} \frac{H(x - s)}{(z - \tau)^{\frac{1}{2}n + \beta}} \left[\exp\left(- \frac{\sigma(x, y; x - s)}{4(z - \tau)} \right) - \exp\left(- \frac{\sigma(x, y; x - s)}{4(y - \tau)} \right) \right] Y\, ds\, d\tau \\ &= J_1 + J_2 + J_3 + J_4. \end{aligned}$$

In J_1 make use of the inequality $(z - y) \geqq (z - \tau)$, in J_2 use the fact that the difference in the brackets for the set $[\alpha, \beta]$ in (24) is less than

$$(2\beta + n) \frac{|z - y|^{\frac{1}{2}}}{(y - \tau)^{\beta+\frac{1}{2}n+\frac{1}{2}}},$$

and in J_3 use the theorem of the mean; then the type of estimate indicated in the inequality (25) is seen to hold for J_1, J_2, and J_3.

Write J_4 in the form

$$J_4 = \int_\eta^y \int_R \frac{H(x - s)}{(z - \tau)^{\frac{1}{2}n+\beta}} \int_y^z \frac{\partial}{\partial v} \exp\left(- \frac{\sigma(x, y; x - s)}{4(v - \tau)}\right) dv \, Y \, ds \, d\tau.$$

Now

$$\left| \int_y^z \frac{\partial}{\partial v} \exp\left(- \frac{\sigma(x, y; x - s)}{4(v - \tau)}\right) dv \right|$$

$$\text{(26)} \qquad \leqq G \int_y^z \frac{-\psi(x - s, v - \tau)}{(v - \tau)} \exp\left(g\psi(x - s, v - \tau)\right) dv$$

$$\leqq (K)G \int_y^z \frac{[-\psi(x - s, v - \tau)]^{\frac{1}{2}}}{v - \tau} \exp\left(h\psi(x - s, v - \tau)\right) dv, \qquad 0 < h < g,$$

where g and G come from property (i) of σ. In the last integral on the right change the variable of integration by letting

$$[-\psi(x - s, v - \tau)]^{\frac{1}{2}} = w.$$

The left member of (26) is then seen to be dominated by

$$(K)G \exp\left(h_1\psi(x - s, z - \tau)\right) |z - y|^{\frac{1}{2}}(y - \tau)^{-\frac{1}{2}}, \quad 0 < h_1 < h.$$

If we make use of the above in J_4, it is easily shown that the type of estimate in (25) applies to J_4, and hence the lemma is proved.

LEMMA 6. *If* $x = (x_1, \cdots, x_n)$, $x' = (x_1, \cdots, x_{i-1}, x_i', x_{i+1}, \cdots, x_n)$ *are two points in* R, *there is a constant* K *such that*

$$|I(x', y) - I(x, y)|$$

$$\text{(27)} \qquad \leqq |x_i' - x_i|^{\frac{1}{2}} \frac{K}{(y - \eta)^{\frac{1}{2}n+1}} \left(\int_\eta^{y'} \int_R |Y| \, ds \, d\tau + \max_{2\tau \geqq y+\eta} |Y(s, \tau)| \right)$$

for $[\alpha, \beta]$ *having any of the values in the set* (24). *In taking the* max $|Y|$ *the* s *are restricted to* $R + R'$.

Consider the case $i = 1$, and assume $x_1' > x_1$. Then break the region R into the regions

R_1 where $s_1 < x_1$,

R_2 where $x_1 \leqq s_1 \leqq x_1'$,

R_3 where $x_1' < s_1$.

The contributions to $I(x', y) - I(x, y)$ from the regions R_1, R_3 are readily shown to satisfy (27). Call J the contribution to $I(x', y) - I(x, y)$ from the region R_2, and write it in the form

$$J = \int_\eta^{\frac{1}{2}(y+\eta)} \int_{R_2} + \int_{\frac{1}{2}(y+\eta)}^{y} \int_{R_2} \frac{1}{(y-\tau)^\beta} [H(x'-s)U(x', y/s, \tau)$$
$$- H(x-s)U(x, y/s, \tau)] \, Y \, ds \, d\tau = J_1 + J_2 .$$

In J_1 use the theorem of the mean and the fact that $y - \tau \geqq \frac{1}{2}(y - \eta)$, and obtain the estimate in (27) concerned with the integral of $|Y|$. To see that J_2 satisfies the part of the inequality (27) concerned with the max $|Y|$, write it as two integrals. Of these two integrals call J_2' the one containing the point x. The integral containing the point x' can be treated as we now treat J_2'.

$$|J_2'| \leqq \max_{2\tau \geqq y+\eta} |Y(s, \tau)| \, (K) \int_{\frac{1}{2}(y+\eta)}^{y} \int_{R_2} \frac{1}{(y-\tau)^v} \exp(h\psi(x-s, y-\tau)) \, ds \, d\tau,$$
$$0 < h < g,$$

where $2v = (2\beta + n - \alpha)$. Making use of the fact that $|x_1 - s_1| \leqq |x_1' - x_1|$ in R_2, we have

$$|J_2'| \leqq |x_1' - x_1|^{\frac{1}{2}} (K) \max_{2\tau \geqq y+\eta} |Y|$$
$$\times \int_{\frac{1}{2}(y+\eta)}^{y} \int_{R_2} \frac{1}{|x_1 - s_1|^{\frac{1}{2}} (y-\tau)^v} \exp(h\psi(x-s, y-\tau)) \, ds \, d\tau.$$

If we remember that $[\alpha, \beta]$ is limited to the set of values in (24), $|J_2'|$ is seen from the above to satisfy the part of the inequality (27) concerned with max $|Y|$. This completes the proof of Lemma 6.

8. **The fundamental solution.** The last two lemmas show that if ξ is in R then $f(x, y; \xi, \eta)$ defined by (21) satisfies a Lipschitz condition in the neighborhood of each point (x, y) where $y > \eta$ and x is in R. It then follows that the function

$$\Gamma(x, y; \xi, \eta) = Z(x, y; \xi, \eta) + \int_\eta^y \int_R Z(x, y; s, \tau) f(s, \tau; \xi, \eta) \, ds \, d\tau$$

satisfies the differential equation $L(u) = 0$ for $y > \eta$ and x, ξ in R. In order to know whether $\Gamma(x, y; \xi, \eta)$ is a fundamental solution of $L(u) = 0$, it remains to show that Γ satisfies property (jj). From the definition of $f(x, y; \xi, \eta)$ and Lemma 3, we see that the integral part of $\Gamma(x, y; \xi, \eta)$ is a function of order $B(\frac{1}{2}(n-1))$. If the integral of a function of order $B(\frac{1}{2}n)$ over a subregion T of R remains finite as y approaches η, then the integral over T of a function of order $B(\frac{1}{2}(n-1))$ will obviously approach zero. Hence

$$(28) \qquad \lim_{y \to \eta} \int_T \varphi(\xi) \Gamma(x, y; \xi, \eta) \, d\xi = \lim_{y \to \eta} \int_T \varphi(\xi) Z(x, y; \xi, \eta) \, d\xi,$$

where $\varphi(\xi)$ is a function continuous in $R + R'$. For x interior to $R - T$ and ξ in T there is a positive constant, say k, such that $\sigma(x, y, x - \xi) > k > 0$, hence the limit in (28) is zero for x interior to $R - T$. By methods similar to those used in the first part of §3, it is readily shown that

$$\lim_{y \to \eta} \int_T \varphi(\xi) Z(x, y; \xi, \eta)\, d\xi = \varphi(x) \qquad (x \text{ interior to } T),$$

hence $\Gamma(x, y; \xi, \eta)$ is a fundamental solution of $L(u) = 0$.

Boundary value problems for $L(u) = 0$ will be considered in a later paper, for now we conclude with our main result.

THEOREM 3. *If the following functions of x, y*

$$\frac{\partial}{\partial y} a_{ij}, \qquad \frac{\partial^2}{\partial x_k \partial x_m} a_{ij}, \qquad a_i, \qquad a \qquad (i, j, k, m = 1, \cdots, n)$$

satisfy a Lipschitz condition of order γ $(0 < \gamma \leqq 1)$ for $\eta \leqq y \leqq y'$ and x in or on the boundary of the bounded region R, and if for these same values of x, y

$$\sum_{i,j=1}^{n} a_{ij} \xi_i \xi_j$$

is a positive definite form, then the parabolic equation

$$\sum_{i,j=1}^{n} a_{ij} \frac{\partial u}{\partial x_i \partial x_j} + \sum_{i=1}^{n} a_i \frac{\partial u}{\partial x_i} + au - \frac{\partial u}{\partial y} = 0$$

has a fundamental solution given by

$$\Gamma(x, y; \xi, \eta) = Z(x, y; \xi, \eta) + \int_\eta^y \int_R Z(x, y; s, \tau) f(s, \tau; \xi, \eta)\, ds\, d\tau.$$

BROWN UNIVERSITY AND DUKE UNIVERSITY.

ON THE ABSOLUTE SUMMABILITY OF FOURIER SERIES, III

By W. C. Randels

In this paper it is proposed to prove the following theorem.

Theorem. *If $f(x)$ is such that at every point y on the closed interval $(-\pi, \pi)$ there are a function $g_y(x)$ and a $\delta > 0$ such that $g_y(x) = f(x)$ for $|x - y| < \delta$ and the Fourier series of $g_y(x)$ is absolutely summable $|C, 1|$, then the Fourier series of $f(x)$ is absolutely summable $|C, 1|$ on $(-\pi, \pi)$.*

This is analogous to a theorem on absolute convergence proved by Wiener.[1]

We must first make some general remarks about absolute summability $|C, 1|$. A series $\sum x_n$ is said to be absolutely summable $|C, 1|$ if

$$\begin{aligned}\sum_{n=1}^{\infty} |\sigma_n^{(1)} - \sigma_{n-1}^{(1)}| &= \sum_{n=1}^{\infty} \left| \frac{1}{n+1} \sum_{\nu=0}^{n} (n - \nu)x_\nu - \frac{1}{n} \sum_{\nu=0}^{n-1} (n - \nu - 1)x_\nu \right| \\ &= \sum_{n=1}^{\infty} \frac{1}{(n+1)n} \left| \sum_{\nu=0}^{n} \nu x_\nu \right| < \infty.\end{aligned} \tag{1}$$

In order to apply this definition to Fourier series in the exponential form we set

$$x_n = (c_n e^{inx} + c_{-n} e^{-inx}).$$

It has been proved[2] that if a series $\sum x_n$ is absolutely summable $|C, 1|$, then

$$\sum_{1}^{\infty} \frac{|x_n|}{n} < \infty.$$

From this it follows that if a Fourier series is absolutely summable $|C, 1|$ over any interval (a, b), then

$$\sum_{-\infty}^{\infty} \frac{|c_n|}{n} < \infty. \tag{2}$$

By the Heine-Borel theorem and the hypotheses of the theorem there will be a finite number of overlapping intervals (δ_i, δ_i') covering $(-\pi, \pi)$ and functions $g_i(x)$ such that the Fourier series of $g_i(x)$ is absolutely summable $|C, 1|$ and $g_i(x) = f(x)$ on (δ_i, δ_i'). These intervals may be chosen so that $\delta_i < \delta_{i-1}' < \delta_{i+1} < \delta_i'$.

Received March 7, 1940. The author wishes to thank Professor N. Wiener for suggesting this problem.

[1] Norbert Wiener, *The Fourier Integral and Certain of Its Applications*, Cambridge University Press, 1933, p. 99, Lemma 6_{15}.

[2] E. Kogbetliantz, Bulletin des Sciences Mathématiques, (2), vol, 49(1925), pp. 234–256.

The functions $h_i(x)$ are now defined by

$$h_i(x) = \begin{cases} A_i(x-\delta_i)^3 + B_i(x-\delta_i)^2, & \delta_i \leqq x < \delta'_{i-1}, \\ 1, & \delta'_{i-1} \leqq x < \delta_{i+1}, \\ 1 - h_{i+1}(x), & \delta_{i+1} \leqq x < \delta'_i, \end{cases}$$

where A_i, B_i are defined by the relations

$$\begin{aligned} 3A_i(\delta'_{i-1} - \delta_i) + 2B_i &= 0, \\ A(\delta'_{i-1} - \delta_i)^3 + B_i(\delta'_{i-1} - \delta_i)^2 &= 1. \end{aligned} \tag{3}$$

The second relation of (3) implies that $h_i(x)$ is continuous, and by the first relation we see that

$$h'_i(\delta_i) = h'_i(\delta'_{i-1}) = h'_i(\delta_{i+1}) = h'_i(\delta'_i) = 0,$$

so that $h'_i(x)$ is also continuous. Therefore

$$\begin{aligned} c_n(h_i) &= \frac{-1}{2\pi} \int_{-\pi}^{\pi} h''_i(x) \frac{e^{inx}}{n^2}\, dx \\ &= O(n^{-3}). \end{aligned} \tag{4}$$

It is also clear that

$$\sum_i h_i(x) = 1,$$

and

$$f(x) = \sum_i g_i(x) h_i(x). \tag{5}$$

The Fourier coefficients of $g_i(x)h_i(x)$ will be given by

$$c_n(g_i \cdot h_i) = \sum_{m=-\infty}^{\infty} c_m(h_i) c_{n-m}(g_i), \tag{6}$$

where the series on the right converges since $h_i(x)$ is of bounded variation. For convenience we put

$$c_n(g_i \cdot h_i) = c_n, \qquad c_n(h_i) = b_n, \qquad c_n(g_i) = a_n.$$

Then from (1) and (6) we must consider

$$\begin{aligned} &\sum_1^{\infty} \frac{1}{n(n+1)} \left| \sum_{\nu=0}^{n} \nu [c_\nu e^{i\nu x} + c_{-\nu} e^{-i\nu x}] \right| \\ &\qquad = \sum_{n=1}^{\infty} \frac{1}{n(n+1)} \left| \sum_{\nu=0}^{n} \nu \sum_{m=-\infty}^{\infty} b_m [a_{\nu-m} e^{i\nu x} + a_{-\nu-m} e^{-i\nu x}] \right| \\ &\qquad \leqq \sum_{n=1}^{\infty} \frac{1}{n(n+1)} \sum_{m=-\infty}^{\infty} | b_m | \left| \sum_{\nu=0}^{n} \nu [a_{\nu-m} e^{i\nu x} + a_{-\nu-m} e^{-i\nu x}] \right| \\ &\qquad = \sum_{m=-\infty}^{\infty} | b_m | \sum_{n=1}^{\infty} \frac{1}{n(n+1)} \left| \sum_{\nu=0}^{n} \nu [a_{\nu-m} e^{i\nu x} + a_{-\nu-m} e^{-i\nu x}] \right|. \end{aligned}$$

If $n \geqq m - 1$, $m > 0$,

$$\begin{aligned}\sum_{\nu=0}^{n} \nu a_{\nu-m} e^{i\nu x} &= \sum_{\nu=0}^{m-1} \nu a_{\nu-m} e^{i\nu x} + \sum_{\nu=m}^{n} \nu a_{\nu-m} e^{i\nu x} \\ &= \sum_{\nu=0}^{m-1} \nu a_{\nu-m} e^{i\nu x} + e^{imx} \sum_{\mu=0}^{n-m} \mu a_\mu e^{i\mu x} + m e^{imx} \sum_{\mu=0}^{n-m} a_\mu e^{i\mu x} \\ &= I_1 + I_2 + I_3,\end{aligned}$$

and

$$\begin{aligned}\sum_{\nu=0}^{n} \nu a_{-\nu-m} e^{-i\nu x} &= \sum_{\nu=0}^{n-2m} \nu a_{-\nu-m} e^{-i\nu x} + \sum_{\nu=n-2m} \nu a_{-\nu-m} e^{-i\nu x} \\ &= \sum_{\nu=n-2m+1}^{n} \nu a_{-\nu-m} e^{-i\nu x} + e^{imx} \sum_{\mu=0}^{n-m} \mu a_{-\mu} e^{-i\mu x} \\ &\qquad - m e^{imx} \sum_{\mu=m}^{n-m} a_{-\mu} e^{-i\mu x} - \sum_{\mu=0}^{m-1} \mu a_{-\mu} e^{-i\mu x} \\ &= H_{1n} + H_{2n} + H_{3n} + H_{4n},\end{aligned}$$

where, if $m > n - m + 1$, H_{3n} drops out and

$$H_{4n} = \sum_{\mu=0}^{n-m} \mu a_{-\mu} e^{-i\mu x}.$$

Then since $\sum a_n$ is absolutely summable $|C, 1|$, we have

$$\sum_{n=m}^{\infty} \frac{1}{n(n+1)} | I_{2n} + H_{2n} | = \sum_{n=m}^{\infty} \frac{1}{n(n+1)} \left| e^{imx} \sum_{\mu=0}^{n-m} \mu \left[a_\mu e^{i\mu x} + a_{-\mu} e^{-i\mu x}\right] \right| < \infty,$$

and

$$\begin{aligned}\sum_{n=m}^{\infty} \frac{1}{n(n+1)} | I_{1n} | &= \sum_{n=m}^{\infty} \frac{1}{n(n+1)} \left| \sum_{\nu=0}^{m-1} a_{\nu-m} e^{i\nu x} \right| \\ &= O\left(m^2 \sum_{n=m}^{\infty} \frac{1}{n(n+1)}\right) = O(m).\end{aligned}$$

Then by (2)

$$\begin{aligned}\sum_{n=m}^{\infty} \frac{1}{n(n+1)} | I_{3n} | &= \sum_{n=m}^{\infty} \frac{1}{n(n+1)} m \left| \sum_{\mu=0}^{n-m} a_\mu e^{i\mu x} \right| \\ &\leqq m \sum_{\mu=0}^{\infty} | a_\mu | \sum_{n=\mu+m}^{\infty} \frac{1}{n(n+1)} \\ &= O\left(m \sum_{\mu=0}^{\infty} \frac{| a_\mu |}{\mu}\right) = O(m).\end{aligned}$$

Similarly,

$$\sum_{n=m}^{\infty} \frac{1}{n(n+1)} | H_{3n} | = \sum_{n=m}^{\infty} \frac{1}{n(n+1)} m \Big| \sum_{\mu=m}^{n-m} a_{-\mu} e^{-i\mu x} \Big|$$
$$= O\left(m \sum_{\mu=m}^{\infty} | a_{-\mu} | \sum_{n=\mu+m}^{\infty} \frac{1}{n(n+1)} \right) = O(m),$$

and, since $(n + m) \leqq 2(n + 1)$, for $n \geqq m$,

$$\sum_{n=m}^{\infty} \frac{1}{n(n+1)} | H_{1n} | = \sum_{n=m}^{\infty} \frac{1}{n(n+1)} \Big| \sum_{\nu=n-2m+1}^{n} \nu a_{-\nu-m} e^{-i\nu x} \Big|$$
$$= O\left(\sum_{n=m}^{\infty} \frac{1}{n+1} \sum_{\nu=n-2m+1}^{n} | a_{-\nu-m} | \right)$$
$$= O\left(\sum_{\mu=0}^{2m-1} \sum_{n=m}^{\infty} \left| \frac{a_{-n-m+\mu}}{n+1} \right| \right) = O(m).$$

Finally

$$\sum_{n=m}^{\infty} \frac{1}{n(n+1)} | H_{4n} | = \sum_{n=m}^{\infty} \frac{1}{n(n+1)} \Big| \sum_{\mu=0}^{m} \mu a_{-\mu} \Big| = O(n),$$

so that

$$\sum_{n=m}^{\infty} \frac{1}{n(n+1)} \Big| \sum_{\nu=0}^{n} \nu [a_{\nu-m} e^{i\nu x} + a_{-\nu-m} e^{-i\nu x}] \Big| = O(m).$$

But we have

$$\sum_{n=0}^{m-1} \frac{1}{n(n+1)} \Big| \sum_{\nu=0}^{n} \nu [a_{\nu-m} e^{i\nu x} + a_{-\nu-m} e^{-i\nu x}] \Big| = O\left(\sum_{n=0}^{m-1} \frac{1}{n(n+1)} n^2 \right) = O(m).$$

Therefore

$$\sum_{n=0}^{\infty} \frac{1}{n(n+1)} \Big| \sum_{\nu=0}^{n} \nu [a_{\nu-m} e^{i\nu x} + a_{-\nu-m} e^{-i\nu x}] \Big| = O(m), \qquad m \geqq 0.$$

The similar result for $m < 0$ can be proved in exactly the same manner, and therefore by (4)

$$\sum_{n=0}^{\infty} \frac{1}{n(n+1)} \Big| \sum_{\nu=0}^{n} \nu [c_\nu e^{i\nu x} + c_{-\nu} e^{-i\nu x}] \Big| = O\left(\sum_{m=-\infty}^{\infty} m b_m \right) = O(1).$$

This proves that the Fourier series of $h_i(x) g_i(x)$ is absolutely summable $| C, 1 |$, and so by (5) $f(x)$, being the sum of a finite number of functions having Fourier series which are absolutely summable $| C, 1 |$, must also have a Fourier series which is absolutely summable $| C, 1 |$. This completes the proof of the theorem.

UNIVERSITY OF OKLAHOMA.

ADDITIVE PRIME NUMBER THEORY IN REAL QUADRATIC FIELDS

By Albert Leon Whiteman

1. **Introduction.** Recent years have seen repeated decisive steps made toward a solution of the Goldbach problem. For almost two centuries this most difficult problem of additive number theory was intractable. Finally, in 1922 Hardy and Littlewood[1] introduced a powerful new method into analysis and proved on the basis of an unproved conjecture about the Riemann zeta-function that every sufficiently large odd number can be represented as the sum of three odd primes and that "almost" every even number is the sum of two primes. In 1930 Schnirelmann,[2] employing an ingenious modification of the Viggo Brun method, proved directly that every even integer can be represented as a sum of not more than 800,000 primes. The number 800,000 was lowered to 2,208 by Romanoff[3] in 1935, to 71 by Heilbronn, Landau and Scherk[4] in 1936, and to 67 by Ricci[5] in 1937. In the same year Vinogradow[6] combined the Hardy-Littlewood method with a new method of his own and gave the first complete proof that every sufficiently large odd number is the sum of three primes. Later, Estermann,[7] extending the ideas initiated by Vinogradow, proved that "almost" every even integer is the sum of two primes.

Less attention has been paid to the problem of representing numbers in an algebraic field as the sum of primes. Indeed, the only contributions in this

Received March 9, 1940.

[1] *Some problems of "Partitio Numerorum", III: On the expression of a number as a sum of primes*, Acta Mathematica, vol. 44(1922), pp. 1–70; *Some problems of "Partitio Numerorum", V: A further contribution to the study of Goldbach's problem*, Proceedings of the London Mathematical Society, (2), vol. 22(1923), pp. 46–56; see also E. Landau, *Vorlesungen über Zahlentheorie*, vol. 1, Leipzig, 1927, pp. 183–234.

[2] *On additive properties of numbers* (Russian with a French résumé), Ann. Inst. Polytechn. Novočerkassk, vol. 14(1930), pp. 3–28.

[3] *On the Goldbach problem* (Russian), Mitt. Forch.-Inst. Math. u. Mech. Univ. Tomsk, vol. 1(1935), pp. 34–38.

[4] *Alle grossen ganzen Zahlen lassen sich als Summe von höchstens 71 Primzahlen darstellen*, Casopis pro Pěstování Matematiky a Fysiky, vol. 65(1936), pp. 117–140.

[5] *Su la congettura di Goldbach e la constante di Schnirelmann*, Annali della Scuola Normale Superiore di Pisa, (2) vol. 6(1937), pp. 91–116.

[6] *Representation of an odd number as a sum of three primes*, Comptes Rendus de l'Académie des Sciences de l'URSS, vol. 15(1937), pp. 169–172.

[7] *On Goldbach's problem: Proof that almost all even positive integers are sums of two primes*, Proceedings of the London Mathematical Society, (2), vol. 44(1938), pp. 307–314; Independent proofs have also been given by J. G. van der Corput, *Sur l'hypothèse de Goldbach pour presque tous les nombres pairs*, Acta Arithmetica, vol. 2(1937), pp. 266–290, and by Tchudakoff, *On the density of the set of even numbers which are not representable as a sum of two primes*, (Russian), Bull. Acad. Sci. de l'URSS, Ser. Math. No. 1, vol. 40(1938), pp. 25–39.

direction are Rademacher's,[8] who, in a series of three memoirs, carried over the Hardy-Littlewood method. He showed that if a certain hypothesis concerning the distribution of the zeros of Hecke's $\zeta(s, \lambda)$-functions is true, every sufficiently "large" "odd" number in an algebraic field is the sum of three primes. In this paper we prove that the same hypothesis implies that "almost" every "even" integer in a real quadratic field is the sum of two primes. We restrict our considerations to a real quadratic field because we wish to avoid numerous cumbersome formulas and because all the difficulties that would be involved in extending our work to a general algebraic field have already been overcome by Rademacher.

The present paper fills a gap in the literature which has existed for a number of years. However, in view of Vinogradow's work, the importance of an investigation of this kind which still makes use of a drastic hypothesis may seem questionable. It should therefore be emphasized that the tools necessary to carry over Vinogradow's intricate analysis have yet to be created. The author believes that the proof given here is a first step toward this goal.

It will be seen in several instances that our methods are very different from those employed by Hardy, Littlewood and Landau. Thus in §4 we consider the analogue of a certain geometric series. An explicit formula for the sum is no longer available and we find it necessary to employ a Fourier expansion in order to estimate the sum. Again in §6 we are able to dispense with an argument concerning Farey dissection which does not seem to have an obvious analogue in algebraic fields.

In conclusion I wish to express my deep appreciation to Professor Hans Rademacher for many valuable suggestions in connection with the preparation of this paper.

2. **Preliminary results.**

Throughout this paper we shall have frequent occasion to use the letter C to denote a positive absolute constant not necessarily the same each time it occurs.

We consider a fixed real quadratic field $k(d^{\frac{1}{2}})$ with positive discriminant d ($d \equiv 0$ or $1 \bmod 4$). The fundamental ideal or "differente"[9] of the field will be denoted by $\mathfrak{d}$. If γ is an algebraic number, then we write $S(\gamma) = \gamma + \gamma'$ and $N(\gamma) = \gamma\gamma'$, where γ' is the conjugate of γ. The ideal $\gamma\mathfrak{d}$ is, in general, fractional. We remove the common ideal factors from its numerator and denominator, put $\gamma\mathfrak{d} = \mathfrak{b}/\mathfrak{a}$, and refer to the ideal $\mathfrak{a}$ as the denominator of $\gamma\mathfrak{d}$.

For future reference we now cite a known theorem[10] on Farey dissection in $k(d^{\frac{1}{2}})$.

[8] *Zur additiven Primzahltheorie algebraischer Zahlkörper*, I, II, III, Abhandlungen aus dem Mathematischen Seminar der Hamburgischen Universität, vol. 3(1924), pp. 109–163 and pp. 331–378; Mathematische Zeitschrift, vol. 27(1926), pp. 321–426. These papers will be cited as Rdm. I, II and III, respectively.

[9] For a definition of the "differente" of an algebraic field and a discussion of its properties see Hecke, *Vorlesungen über die Theorie der algebraischen Zahlen*, Leipzig, 1923, pp. 131–136.

[10] Rdm. III, p. 329, Lemma 4.

Let $\mathfrak{M}$ denote the totality of all pairs of real numbers (z, z'). Let M be a positive integer such that $M > N(\mathfrak{d})$.[11] Then the Farey dissection of $\mathfrak{M}$ with respect to M is described by

Lemma 1. *The manifold $\mathfrak{M}$ can be divided into non-overlapping rectifiable regions $\mathfrak{F}_\gamma$ which completely cover the plane. The regions $\mathfrak{F}_\gamma$ correspond to real quadratic numbers γ such that $\gamma\mathfrak{d}$ has an ideal denominator $\mathfrak{a}$ whose norm does not exceed M. Furthermore, we have*

(a) *The region $\mathfrak{F}_\gamma$ lies entirely within the region defined by the inequality*

$$(2.11)\qquad (1 + M\,|\,z - \gamma\,|)(1 + M\,|\,z' - \gamma'\,|) \leqq \frac{4M}{N(\mathfrak{a})}.$$

(b) *The region $\mathfrak{F}_\gamma$ contains the region defined by the inequality*

$$(2.12)\qquad (1 + M\,|\,z - \gamma\,|)(1 + M\,|\,z' - \gamma'\,|) \leqq \frac{M}{[N(\mathfrak{d})]^{\frac{1}{2}} N(\mathfrak{a})}.$$

(c) *If $\gamma_1 \equiv \gamma_2$ (mod $1/\mathfrak{d}$), then $\mathfrak{F}_{\gamma_1}$ and $\mathfrak{F}_{\gamma_2}$ are congruent and homothetic. In particular, if γ runs through an incongruent residue system* mod $1/\mathfrak{d}$, *then the aggregate of the corresponding $\mathfrak{F}_\gamma$ forms a region $\mathfrak{B}$ having the following property: corresponding to every point (z, z') there is just one point that is congruent* mod $1/\mathfrak{d}$ *and lies within $\mathfrak{B}$.*

We shall also need the following lemma.

Lemma 2. *For (z, z') in $\mathfrak{F}_\gamma$ we have*

$$|\,z - \gamma\,| + |\,z' - \gamma'\,| \leqq 2, \qquad |\,z - \gamma\,|\,|\,z' - \gamma'\,| < \frac{1}{d}.$$

This lemma follows at once from Rademacher's theory of Farey dissection in an algebraic field.[12] Let Ω_1, Ω_2 be a basis of $k(d^{\frac{1}{2}})$. Then we have for (z, z') in $\mathfrak{F}_\gamma$

$$|\,z - \gamma\,| \leqq \frac{M^{-\frac{1}{2}}}{|\,b_1\Omega_1 + b_2\Omega_2\,|}, \qquad |\,z' - \gamma'\,| \leqq \frac{M^{-\frac{1}{2}}}{|\,b_1\Omega_1' + b_2\Omega_2'\,|},$$

where

$$0 < |\,b_1\Omega_1 + b_2\Omega_2\,| \leqq M^{\frac{1}{2}}, \qquad 0 < |\,b_1\Omega_1' + b_2\Omega_2'\,| \leqq M^{\frac{1}{2}},$$

$$|\,b_1\Omega_1 + b_2\Omega_2\,|\,|\,b_1\Omega_1' + b_2\Omega_2'\,| \geqq N(\mathfrak{a}).$$

[11] It should be mentioned that the "differente" plays an important rôle in the investigation of the fundamental generating function (2.22); for this reason it was found convenient to introduce it into the theory of Farey dissection.

[12] Rdm. III, pp. 327–332.

We next recall that $N(\mathfrak{d}) = d$.[13] Hence

$$|z - \gamma| \leqq \frac{M^{-\frac{1}{2}}}{|b_1\Omega_1 + b_2\Omega_2|} \leqq \frac{|b_1\Omega_1' + b_2\Omega_2'|M^{-\frac{1}{2}}}{N(\mathfrak{a})} \leqq \frac{1}{N(\mathfrak{a})} \leqq 1,$$

$$|z' - \gamma'| \leqq \frac{M^{-\frac{1}{2}}}{|b_1\Omega_1' + b_2\Omega_2'|} \leqq \frac{|b_1\Omega_1 + b_2\Omega_2|M^{-\frac{1}{2}}}{N(\mathfrak{a})} \leqq \frac{1}{N(\mathfrak{a})} \leqq 1,$$

$$|z - \gamma||z' - \gamma'| \leqq \frac{1}{|b_1\Omega_1 + b_2\Omega_2||b_1\Omega_1' + b_2\Omega_2'|M} \leqq \frac{1}{N(\mathfrak{a})M} < \frac{1}{d}.$$

This completes the proof of the lemma.

The fundamental function employed by Rademacher in his investigations on real quadratic fields is the following series in two complex variables t and t':

$$\sum \exp(-\omega t - \omega' t'), \qquad \omega \succ 0, \tag{2.21}$$

where the notation $\omega \succ 0$ means that ω runs through all the totally positive[14] primes[15] in the field. But, as will be explained later, this function leads to a result which is not suitable for our present purposes. It was found more convenient to investigate the function

$$f(t, t') = \sum \log N(\omega) \exp(-\omega t - \omega' t'), \qquad \omega \succ 0, \tag{2.22}$$

where

$$t = w - 2\pi i z, \qquad t' = w' - 2\pi i z', \tag{2.23}$$

$$w, w' > 0; \qquad z, z' \text{ real}. \tag{2.24}$$

The conditions in (2.23) and (2.24) assure the convergence of the series in (2.22) and indeed the uniform convergence for $w, w' \geqq R > 0$.[16] Hence, for m a positive integer, we clearly have

$$f^m(t, t') = \sum A_m(\mu) \exp(-\mu t - \mu' t'), \qquad \mu \succ 0, \tag{2.25}$$

where μ runs through all the totally positive integers in the field and

$$\begin{gathered} A_m(\nu) = \sum \log N(\omega_1) \log N(\omega_2) \cdots \log N(\omega_m), \\ \omega_1, \omega_2, \cdots, \omega_m \succ 0, \qquad \omega_1 + \omega_2 + \cdots + \omega_m = \nu. \end{gathered} \tag{2.26}$$

In part I of the author's doctoral dissertation at the University of Pennsylvania, Rademacher's method was applied to (2.22) and (2.26) and the following two principal results were obtained.

[13] Hecke, loc. cit., p. 133.

[14] An integer μ in a real quadratic field is said to be totally positive if $\mu > 0$, $\mu' > 0$.

[15] An algebraic integer is said to be prime if the principal ideal which it generates is a prime ideal.

[16] Rdm. I, pp. 112–113.

LEMMA 3. *Let θ denote*[17] *the upper bound of the real parts of all the zeros of all the Hecke $\zeta(s, \lambda)$-functions*[18] *in a real quadratic field. Let d denote the discriminant, h the class number, and η the fundamental unit of the field. Then we have*

$$\left| f(w - 2\pi iz, w' - 2\pi iz') - \frac{\mu(\mathfrak{a})}{2h \log \eta\, \varphi(\mathfrak{a})} \frac{1}{(w - 2\pi i(z - \gamma))(w' - 2\pi i(z' - \gamma'))} \right|$$

$$< CN(\mathfrak{a})^{\frac{1}{2}+2\epsilon} \frac{(1 + 2\pi |z - \gamma|/w)^{\frac{1}{2}+\epsilon}(1 + 2\pi |z' - \gamma'|/w')^{\frac{1}{2}+\epsilon}}{w^{\theta+\epsilon} w'^{\theta+\epsilon}} \tag{2.3}$$

$$+ CN(\mathfrak{a})^{1+\epsilon} \frac{(1 + 2\pi |z - \gamma|/w)^{\frac{1}{2}+\epsilon}(1 + 2\pi |z' - \gamma'|/w')^{\frac{1}{2}+\epsilon}}{w^{\frac{1}{2}} w'^{\frac{1}{2}}},$$

where $f(w - 2\pi iz, w' - 2\pi iz')$ is defined by (2.22), (2.23), *and* (2.24). *Here $\mu(\mathfrak{a})$ and $\varphi(\mathfrak{a})$ denote as usual the Möbius μ-function and the Euler φ-function respectively. The ideal $\mathfrak{a}$ is the denominator of $\gamma\mathfrak{d}$.*

LEMMA 4. *Let $A_m(\nu)$ be defined by* (2.26). *Then, with the notation of Lemma* 3, *we have for $m \geqq 3$*

$$A_m(\nu) = \frac{d^{\frac{1}{2}}}{(2h \log \eta)^m \Gamma^2(m)} N(\nu)^{m-1} \mathfrak{S}_m(\nu) + O(N(\nu)^{m-1+(\theta-\frac{1}{2})+\epsilon}), \tag{2.41}$$[19]

where $\mathfrak{S}_m(\nu)$ is the so-called "singular series"

$$\mathfrak{S}_m(\nu) = \sum \left(\frac{\mu(\mathfrak{a})}{\varphi(\mathfrak{a})}\right)^m \sum_{\gamma} e^{-2\pi i S(\gamma\nu)}, \qquad \gamma \bmod \frac{1}{\mathfrak{d}} \tag{2.42}$$

$$= \prod_{\mathfrak{P} \nmid \nu} \left(1 - \left(\frac{-1}{N(\mathfrak{P}) - 1}\right)^m\right) \prod_{\mathfrak{P} | \nu} \left(1 + \frac{(-1)^m}{(N(\mathfrak{P}) - 1)^{m-1}}\right). \tag{2.43}$$

We now prove a lemma which will be very useful in the sequel.

LEMMA 5. *Let*

$$g(w, z) = \sum_{\mu} a_\mu \exp\left[-(\mu w + \mu' w') + 2\pi i(\mu z + \mu' z')\right], \tag{2.51}$$

where $a_\mu = O(|N(\mu)|^\epsilon)$, $\mu > 0$ and w, z are taken as in (2.24). *Let ω_1, ω_2 be a basis of the fractional ideal $1/\mathfrak{d}$. Put*

$$z = \omega_1 x_1 + \omega_2 x_2, \quad z' = \omega_1' x_1 + \omega_2' x_2, \qquad x_1, x_2 \text{ real}. \tag{2.52}$$

[17] The exact value of θ has not yet been determined. However, it is known that $\frac{1}{2} \leqq \theta \leqq 1$.

[18] For the definition of the Hecke $\zeta(s, \lambda)$-functions and a discussion of their properties see Rdm. III, pp. 362–367.

[19] If we had chosen (2.21) instead of (2.23) as our fundamental generating function, then the result corresponding to (2.41) would have contained a troublesome factor $(\log N(\nu))^m$ in the denominator of the principal term.

Then we have

$$\int_0^1\int_0^1 |g(w,z)|^2\,dx_1\,dx_2 = \sum_\mu |a_\mu|^2 e^{-2\mu w-2\mu' w'}, \qquad \mu \succ 0. \tag{2.53}$$

Rademacher has established the convergence of the series in (2.51) and indeed the uniform convergence for $w,\,w' \geqq R > 0$.[20] Let $\bar{a}$ be the conjugate of a. Then

$$\int_0^1\int_0^1 |g(w,z)|^2\,dx_1\,dx_2 = \int_0^1\int_0^1 g(w,z)\overline{g(w,z)}\,dx_1\,dx_2$$

$$= \int_0^1\int_0^1 \sum_\mu a_\mu \exp\,[-(\mu w + \mu' w') + 2\pi i(\mu z + \mu' z')]$$

$$\cdot \sum_\nu \bar{a}_\nu \exp\,[-(\nu w + \nu' w') - 2\pi i(\nu z + \nu' z')]\,dx_1\,dx_2$$

$$= \int_0^1\int_0^1 \sum_\mu\sum_\nu a_\mu \bar{a}_\nu \exp\,[-(\mu+\nu)w - (\mu'+\nu')w']$$

$$\cdot \exp\,[2\pi i x_1 S((\mu-\nu)\omega_1)]\exp\,[2\pi i x_2 S((\mu-\nu)\omega_2)]\,dx_1\,dx_2$$

$$= \sum_\mu\sum_\nu a_\mu \bar{a}_\nu \exp\,[-(\mu+\nu)w - (\mu'+\nu')w']$$

$$\cdot \int_0^1 \exp\,[2\pi i x_1 S((\mu-\nu)\omega_1)]\,dx_1 \int_0^1 \exp\,[2\pi i x_2 S((\mu-\nu)\omega_2)]\,dx_2,$$

where $\mu \succ 0$, $\nu \succ 0$. Since the two equations

$$S((\mu-\nu)\omega_1) = 0, \qquad S((\mu-\nu)\omega_2) = 0$$

have simultaneous solutions if and only if $\mu = \nu$, and since $S((\mu-\nu)\omega_1)$ and $S((\mu-\nu)\omega_2)$ are rational integers, the lemma now follows immediately.

3. **The Farey dissection.** Since Lemma 4 is not necessarily true for $m = 2$, we are led to investigate the average deviation of $A_m(\nu)$ with respect to its asymptotic representation. Accordingly we put

$$D_m(V, V') = \sum \left(A_m(\mu) - \frac{d^{\frac{1}{2}}}{(2h\log\eta)^m\Gamma^2(m)} N(\mu)^{m-1}\mathfrak{S}_m(\mu)\right)^2, \tag{3.11}$$

where now $m \geqq 2$ and the sum is taken over all totally positive integers μ lying in the rectangle $0 < \mu \leqq V$, $0 < \mu' \leqq V'$ and V, V' denote two arbitrary positive real numbers which are not necessarily the conjugates of an algebraic number. The convergence of $\mathfrak{S}_m(\mu)$ for $m = 2$ follows directly from its representation in (2.43) as an infinite product. In order to study (3.11) by means of the

[20] Rdm. I, pp. 112-113.

Hardy-Littlewood method we first apply Lemma 5 and obtain after a slight modification

$$
\begin{aligned}
D_m(V, V') &\leqq e^{2Vw+2V'w'} \sum_{\mu} \Bigg(A_m(\mu) - \frac{d^{\frac{1}{2}}}{(2h \log \eta)^m \Gamma^2(m)} \\
&\qquad\qquad \cdot N(\mu)^{m-1} \mathfrak{S}_m(\mu) \Bigg)^2 e^{-2\mu w - 2\mu' w'} \\
&= e^{2Vw+2V'w'} \int_0^1 \int_0^1 \left| \sum_{\mu} \left(A_m(\mu) - \frac{d^{\frac{1}{2}}}{(2h \log \eta)^m \Gamma^2(m)} N(\mu)^{m-1} \mathfrak{S}_m(\mu) \right) \right. \\
&\qquad\qquad \left. \cdot \exp\left[-(\mu w + \mu' w') + 2\pi i(\mu z + \mu' z') \right] \right|^2 dx_1\, dx_2, \qquad \mu \succ 0.
\end{aligned} \tag{3.12}
$$

For the sake of brevity we now introduce the abbreviation

$$
\begin{aligned}
F_m(w, z) = \frac{d^{\frac{1}{2}}}{(2h \log \eta)^m \Gamma^2(m)} &\sum_{\mu} N(\mu)^{m-1} \mathfrak{S}_m(\mu) \\
&\cdot \exp\left[-(\mu w + \mu' w') + 2\pi i(\mu z + \mu' z') \right], \qquad \mu \succ 0.
\end{aligned} \tag{3.13}
$$

Then, substituting from (2.23), (2.25) and (3.13) into (3.12), we have

$$
D_m(V, V') \leqq \exp\,(2Vw + 2V'w') \int_0^1 \int_0^1 | f^m(w, z) - F_m(w, z) |^2\, dx_1\, dx_2 . \tag{3.14}
$$

At this point it is convenient to assume that

$$
\eta_1^{-1} \leqq \frac{V'}{V} \leqq \eta_1 , \tag{3.21}
$$

where η_1 denotes the fundamental unit mod 1.[21] (Later we shall see that this restriction does not entail any loss of generality.) Furthermore, we shall henceforth take

$$
w = \frac{1}{V}, \qquad w' = \frac{1}{V'}. \tag{3.22}
$$

We are now in the position to introduce the Farey dissection described in Lemma 1. To accomplish this we observe that the transformation (2.52) which was used in obtaining (3.12) carries the region of integration in the variables x_1, x_2 into a new region of integration in the variables z, z'. This new region of integration is precisely the region $\mathfrak{B}$ described in Lemma 1. On the other hand, the transformation (2.52) carries all pairs of real numbers (z, z') into all

[21] In general we let $\eta(\mathfrak{a}) > 1$ denote the fundamental unit mod $\mathfrak{a}$, that is, $\eta(\mathfrak{a})$ is totally positive and $\eta(\mathfrak{a}) \equiv 1 \pmod{\mathfrak{a}}$; moreover, every other unit with the same property can be represented as $\eta(\mathfrak{a})^n$ where n is a rational integer. Hence $\eta_1 = \eta$ or $\eta_1 = \eta^2$ according as the fundamental unit η is totally positive or not. Another name for η_1 is the totally positive fundamental unit.

pairs of real numbers (x_1, x_2). Therefore, each $\mathfrak{F}_\gamma$ is carried over into $\mathfrak{F}'_\gamma$, say. In this manner we obtain from (3.14) that

$$D_m(V, V') \leqq C \sum \iint_{\mathfrak{F}'_\gamma} | f^m(w, z) - F_m(w, z) |^2 \, dx_1 \, dx_2, \tag{3.3}$$

$$\gamma \bmod 1/\mathfrak{d}, \qquad N(\mathfrak{a}) \leqq M.$$

We now return to Lemma 3. In order to apply the Farey dissection we put

$$z = \gamma + y, \qquad z' = \gamma' + y', \tag{3.41}$$

and

$$M = (VV')^{\frac{1}{2}}. \tag{3.42}$$

Then, by (2.11), (2.12) and (3.41) we have for $\gamma + y$ in $\mathfrak{F}_\gamma$

$$(1 + M | y |)(1 + M | y' |) < C \frac{M}{N(\mathfrak{a})}. \tag{3.51}$$

Note next that (3.21) clearly implies that

$$C(VV')^{\frac{1}{4}} < \frac{V}{V'} < C(VV')^{\frac{1}{4}}. \tag{3.52}$$

From this it follows, in combination with (3.22) and (3.42), that for $\gamma + y$ in $\mathfrak{F}_\gamma$ we have

$$N(\mathfrak{a}) \left(1 + \frac{2\pi | y |}{w}\right)\left(1 + \frac{2\pi | y' |}{w'}\right) \tag{3.53}$$

$$< N(\mathfrak{a})(1 + 2\pi CM | y |)(1 + 2\pi CM | y' |) < C(VV')^{\frac{1}{4}}.$$

Applying this result on (2.3) and introducing the substitutions (3.22) and (3.41) give

$$\left| f\left(\frac{1}{V}, \gamma + y\right) - \frac{\mu(\mathfrak{a})}{2h \log \eta \, \varphi(\mathfrak{a})} \frac{1}{(w - 2\pi i y)(w' - 2\pi i y')} \right| \tag{3.61}$$

$$< CN(\mathfrak{a})^{\epsilon}(VV')^{\theta + \frac{1}{4} + \frac{1}{2}\epsilon} + CN(\mathfrak{a})^{\frac{1}{2}}(VV')^{\frac{1}{2} + \frac{1}{4} + \frac{1}{2}\epsilon},$$

for $\gamma + y$ within $\mathfrak{F}_\gamma$. Since $\frac{1}{2} \leqq \theta \leqq 1$ and since $N(\mathfrak{a}) \leqq M$, it follows finally that for $\gamma + y$ in $\mathfrak{F}_\gamma$ we have

$$\left| f\left(\frac{1}{V}, \gamma + y\right) - \frac{\mu(\mathfrak{a})}{2h \log \eta \, \varphi(\mathfrak{a})} \frac{1}{(w - 2\pi i y)(w' - 2\pi i y')} \right| < C(VV')^{\theta + \frac{1}{4} + 2\epsilon}. \tag{3.62}$$

4. **A Fourier expansion.** We now return to the function $F_m(w, z)$ defined in (3.13). Introducing the substitution (2.23) and making use of (2.42), we get

$$F_m(w, z) = \frac{d^{\frac{1}{2}}}{(2h \log \eta)^m \Gamma^2(m)} \sum_{\mu} N(\mu)^{m-1} \sum A_m(\mu; \mathfrak{a}) \exp(-\mu t - \mu' t'), \tag{4.11}$$

$$\mu \succ 0,$$

where $\mathfrak{a}$ is the summand in the second summation, and where, for the sake of brevity, we have put

$$(4.12)\qquad A_m(\mu;\mathfrak{a}) = \left(\frac{\mu(\mathfrak{a})}{\varphi(\mathfrak{a})}\right)^m \sum_\gamma \exp\left[-2\pi i S(\gamma\mu)\right], \qquad \gamma \bmod \frac{1}{\mathfrak{d}}.$$

The absolute convergence of the last sum follows from the well-known formula[22]

$$(4.13)\qquad \sum_\gamma \exp\left[-2\pi i S(\gamma\mu)\right] = \sum N(\mathfrak{b})\mu(\mathfrak{a}/\mathfrak{b}),$$

where the summation is for γ mod $1/\mathfrak{d}$ and $\mathfrak{b} \mid (\mathfrak{a}, \mu)$. From this formula we easily deduce an estimate independent of $\mathfrak{a}$, namely,

$$\left|\sum_\gamma \exp\left[-2\pi i S(\gamma\mu)\right]\right| \leqq \sum N(\mathfrak{b}) \leqq CN(\mu)^2,$$

where the summation is for γ mod $1/\mathfrak{d}$ and $\mathfrak{b} \mid \mu$. Because of the absolute convergence we may now interchange the signs of summation in (4.11). We get thus

$$(4.14)\qquad \begin{aligned} F_m(w, z) &= \frac{d^{\frac{1}{2}}}{(2h\log\eta)^m\Gamma^2(m)} \sum\sum N(\mu)^{m-1} A_m(\mu;\mathfrak{a}) \exp\left(-\mu t - \mu' t'\right) \\ &= \frac{d^{\frac{1}{2}}}{(2h\log\eta)^m\Gamma^2(m)} \sum\sum N(\mu)^{m-1}\left(\frac{\mu(\mathfrak{a})}{\varphi(\mathfrak{a})}\right)^m \\ &\qquad\qquad \cdot\sum_\gamma \exp\left[-2\pi i S(\gamma\mu)\right]\exp\left(-\mu t - \mu' t'\right) \\ &= \frac{d^{\frac{1}{2}}}{(2h\log\eta)^m\Gamma^2(m)} \sum\left(\frac{\mu(\mathfrak{a})}{\varphi(\mathfrak{a})}\right)^m \sum_\gamma\sum_\mu N(\mu)^{m-1}\exp\left(-\mu u - \mu' u'\right), \end{aligned}$$

where we have put

$$(4.15)\qquad u = w - 2\pi i(z - \gamma), \qquad u' = w' - 2\pi i(z' - \gamma'),$$

and where the summation is for $\mathfrak{a}$; γ mod $1/\mathfrak{d}$; and $\mu > 0$.

Our next object is to estimate the function

$$(4.21)\qquad g(u, u') = \sum_\mu N(\mu)^{m-1}\exp\left(-\mu u - \mu' u'\right), \qquad \mu > 0,$$

which appears in (4.14). In order to accomplish this we first expand $g(u, u')$ into a Fourier series by means of the well-known Poisson method. Let Ω_1, Ω_2 be a basis of $k(d^{\frac{1}{2}})$. Then the function

$$(4.22)\qquad G(x, y) = \sum (\mu + \xi)^{m-1}(\mu' + \xi')^{m-1}\exp\left[-(\mu + \xi)u - (\mu' + \xi')u'\right],$$

where

$$(4.23)\qquad \xi = x\Omega_1 + y\Omega_2, \qquad \xi' = x\Omega_1' + y\Omega_2', \qquad x, y \text{ real},$$

and the summation extends over all algebraic integers μ for which $\mu + \xi > 0$, $\mu' + \xi' > 0$, is periodic in both x and y with the period 1.[23] It is not difficult

[22] Rdm. III, p. 335.

[23] This ingenious application of the Poisson method is due to Siegel, Mathematische Annalen, vol. 87(1922), pp. 5-9.

to show that it is legitimate to develop $G(x, y)$ into an absolutely convergent Fourier series of the form[24]

$$G(x, y) = \sum_{k=-\infty}^{\infty} \sum_{l=-\infty}^{\infty} A_{kl} \exp [2\pi i(kx + ly)], \tag{4.24}$$

with

$$\begin{aligned} A_{kl} = \int_0^1 \int_0^1 \sum_\mu \epsilon_{\mu+\xi}\epsilon_{\mu'+\xi'}(\mu + \xi)^{m-1}(\mu' + \xi')^{m-1} \\ \cdot\exp [-(\mu + \xi)u - (\mu' + \xi')u' - 2\pi i(kx + ly)]\, dx\, dy, \end{aligned} \tag{4.25}$$

where the sum extends over all algebraic integers μ and $\epsilon_a = 0$ or 1 according as $a < 0$ or $a \geqq 0$. From (4.23) we have

$$\left|\frac{\partial(\xi, \xi')}{\partial(x, y)}\right| = \begin{vmatrix} \Omega_1 & \Omega_1' \\ \Omega_2 & \Omega_2' \end{vmatrix} = d^{\frac{1}{2}} \tag{4.26}$$

and

$$\begin{aligned} \exp [-2\pi i(kx + ly)] = \exp [-2\pi i d^{-\frac{1}{2}}\xi(k\Omega_2' - l\Omega_1') \\ - 2\pi i d^{-\frac{1}{2}}\xi'(-k\Omega_2 + l\Omega_1)]. \end{aligned} \tag{4.27}$$

If μ is an algebraic integer, then $S(\mu d^{-\frac{1}{2}})$ is a rational integer. We may therefore write the last equation in the form

$$\begin{aligned} \exp [-2\pi i(kx + ly)] = \exp \{-2\pi i d^{-\frac{1}{2}}[(\mu + \xi)(k\Omega_2' - l\Omega_1')] \\ - 2\pi i d^{-\frac{1}{2}}[(\mu' + \xi')(-k\Omega_2 + l\Omega_1)]\}. \end{aligned} \tag{4.28}$$

Introducing the substitutions (4.23), (4.26), (4.28) into (4.25) we find

$$\begin{aligned} A_{kl} = d^{-\frac{1}{2}} \sum_\mu \int\int \epsilon_{\mu+\xi}\epsilon_{\mu'+\xi'}(\mu + \xi)^{m-1}(\mu' + \xi')^{m-1} \\ \cdot\exp \{-(\mu + \xi)[u + 2\pi i d^{-\frac{1}{2}}(k\Omega_2' - l\Omega_1')]\} \\ \cdot\exp \{-(\mu' + \xi')[u' + 2\pi i d^{-\frac{1}{2}}(-k\Omega_2 + l\Omega_1)]\}\, d\xi\, d\xi', \end{aligned}$$

where the double integral extends over the image of the unit square in the xy-plane on the $\xi\xi'$-plane. We now set $\mu + \xi = U$, $\mu' + \xi' = U'$ and obtain

$$\begin{aligned} A_{kl} &= d^{-\frac{1}{2}} \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \epsilon_U\epsilon_{U'}(UU')^{m-1} \exp \{-U[u + 2\pi i d^{-\frac{1}{2}}(k\Omega_2' - l\Omega_1')]\} \\ &\qquad \cdot\exp \{-U'[u' + 2\pi i d^{-\frac{1}{2}}(-k\Omega_2 + l\Omega_1)]\}\, dU\, dU' \\ &= d^{-\frac{1}{2}} \int_0^{\infty} U^{m-1} \exp \{-U[u + 2\pi i d^{-\frac{1}{2}}(k\Omega_2' - l\Omega_1')]\}\, dU \\ &\qquad \cdot\int_0^{\infty} U'^{m-1} \exp \{-U'[u' + 2\pi i d^{-\frac{1}{2}}(-k\Omega_2 + l\Omega_1)]\}\, dU'. \end{aligned}$$

[24] Compare Siegel, loc. cit., pp. 5–9.

The last two integrals contain complex parameters. However, absolute convergence is assured since $\mathfrak{R}(u) > 0$, $\mathfrak{R}(u') > 0$. Therefore, we have finally

$$(4.29)\qquad A_{kl} = d^{-\frac{1}{2}}\Gamma^2(m)\,\frac{1}{[u + 2\pi i d^{-\frac{1}{2}}(k\Omega_2' - l\Omega_1')]^m[u' + 2\pi i d^{-\frac{1}{2}}(-k\Omega_2 + l\Omega_1)]^m}.$$

Introducing (4.27) and (4.29) into (4.24) we get

$$G(x, y) = \sum_{k=-\infty}^{\infty}\sum_{l=-\infty}^{\infty} d^{-\frac{1}{2}}\Gamma^2(m)$$

$$\cdot\frac{\exp\,[2\pi i d^{-\frac{1}{2}}\xi(k\Omega_2' - l\Omega_1') + 2\pi i d^{-\frac{1}{2}}\xi'(-k\Omega_2 + l\Omega_1)]}{[u + 2\pi i d^{-\frac{1}{2}}(k\Omega_2' - l\Omega_1')]^m[u' + 2\pi i d^{-\frac{1}{2}}(-k\Omega_2 + l\Omega_1)]^m}.$$

Since Ω_2', $-\Omega_1'$ is also a basis of $k(d^{\frac{1}{2}})$, the last equation may be written in the simpler form

$$\sum_{\substack{\mu+\xi>0\\ \mu'+\xi'>0}} (\mu + \xi)^{m-1}(\mu' + \xi')^{m-1}\exp\,[-(\mu + \xi)u - (\mu' + \xi')u']$$

$$= d^{-\frac{1}{2}}\Gamma^2(m)\sum_{\lambda}\frac{\exp\,[2\pi i d^{-\frac{1}{2}}(\xi\lambda - \xi'\lambda')]}{(u + 2\pi i d^{-\frac{1}{2}}\lambda)^m(u' - 2\pi i d^{-\frac{1}{2}}\lambda')^m},$$

where λ runs through all the algebraic integers in $k(d^{\frac{1}{2}})$. For our application we put $x = 0$, $y = 0$ and obtain the following transformation of (4.21)

$$(4.31)\qquad \begin{aligned} g(u, u') &= d^{-\frac{1}{2}}\Gamma^2(m)\sum_{\lambda}\frac{1}{(u + 2\pi i d^{-\frac{1}{2}}\lambda)^m(u' - 2\pi i d^{-\frac{1}{2}}\lambda')^m}\\ &= d^{-\frac{1}{2}}\frac{\Gamma^2(m)}{(uu')^m} + \sum_{\lambda\neq 0}\frac{1}{(u + 2\pi i d^{-\frac{1}{2}}\lambda)^m(u' - 2\pi i d^{-\frac{1}{2}}\lambda')^m}.\end{aligned}$$

The first term in (4.31) will be used as an approximation of $g(u, u')$ and the second term will be used to estimate the resulting error. From (3.22), (3.41) and (4.15) we have

$$(4.32)\qquad \begin{aligned} &\sum_{\lambda\neq 0}\frac{1}{|u + 2\pi i d^{-\frac{1}{2}}\lambda|^m|u' - 2\pi i d^{-\frac{1}{2}}\lambda'|^m}\\ &\qquad < C\sum_{\lambda\neq 0}\frac{1}{(V^{-1} + 2\pi|y - d^{-\frac{1}{2}}\lambda|)^m(V'^{-1} + 2\pi|y' + d^{-\frac{1}{2}}\lambda'|)^m}.\end{aligned}$$

In order to estimate the last sum we first show that for $\gamma + y$ in $\mathfrak{F}_\gamma$ and for $\lambda \neq 0$ we have

$$(4.33)\qquad \max\,(|y - d^{-\frac{1}{2}}\lambda|, |y' + d^{-\frac{1}{2}}\lambda'|) > C.$$

This can be proved in the following way. By Lemma 2, we have for $\gamma + y$ in $\mathfrak{F}_\gamma$

$$|y - d^{-\frac{1}{2}}\lambda| + |y' + d^{-\frac{1}{2}}\lambda'| \geqq |d^{-\frac{1}{2}}\lambda| + |d^{-\frac{1}{2}}\lambda'| - |y| - |y'|$$

$$\geqq d^{-\frac{1}{2}}(|\lambda| + |\lambda'|) - 2.$$

Therefore, for $|\lambda| + |\lambda'| \geqq 3d^{\frac{1}{2}}$ it follows that $\max(|y - d^{-\frac{1}{2}}\lambda|, |y' + d^{-\frac{1}{2}}\lambda'|) > \frac{1}{2}$. Let us now observe that the two quantities $|y - d^{-\frac{1}{2}}\lambda|$ and $|y' + d^{-\frac{1}{2}}\lambda'|$ cannot simultaneously equal zero. Otherwise we would have

$$|yy'| = |d^{-\frac{1}{2}}\lambda \cdot d^{-\frac{1}{2}}\lambda'| \geqq \frac{1}{d},$$

and this contradicts Lemma 2. Since there are only a finite number of algebraic integers λ for which $|\lambda| + |\lambda'| < 3d^{\frac{1}{2}}$, it is now clear that the inequality (4.33) holds for all λ.

Next, a simple geometrical consideration shows that the number of solutions of the inequalities

$$k \leqq |y - d^{-\frac{1}{2}}\lambda| < k + 1, \qquad l \leqq |y' + d^{-\frac{1}{2}}\lambda'| < l + 1 \tag{4.4}$$

in integers λ, λ' is less than C.

Applying (4.33) and (4.4) to the right member of (4.32), we obtain finally for $\gamma + y$ in $\mathfrak{F}_\gamma$

$$\begin{aligned}\sum_{\lambda \neq 0} &\frac{1}{(V^{-1} + 2\pi|y - d^{-\frac{1}{2}}\lambda|)^m(V'^{-1} + 2\pi|y' + d^{-\frac{1}{2}}\lambda'|)^m} \\ &\leqq C\left\{\frac{1}{(2\pi CV^{-1})^m} + \frac{1}{(2\pi CV'^{-1})^m} + \sum_{k=1}^{\infty}\frac{1}{(2\pi kV'^{-1})^m}\right. \\ &\qquad \left. + \sum_{l=1}^{\infty}\frac{1}{(2\pi lV^{-1})^m} + \sum_{k=1}^{\infty}\sum_{l=1}^{\infty}\frac{1}{(4\pi^2 kl)^m}\right\} \\ &\leqq C(VV')^{\frac{1}{2}m},\end{aligned} \tag{4.5}$$

where we have used (3.52) in the last step. Combining (4.31) and (4.32) with (4.5) now yields the desired estimate

$$\left|g(u, u') - d^{-\frac{1}{2}}\Gamma^2(m)\frac{1}{(uu')^m}\right| < C(VV')^{\frac{1}{2}m}, \tag{4.6}$$

for $\gamma + y$ in $\mathfrak{F}_\gamma$. Let us put

$$F_\gamma(w, z) = \frac{d^{\frac{1}{2}}}{(2h \log \eta)^m \Gamma^2(m)}\left(\frac{\mu(\mathfrak{a})}{\varphi(\mathfrak{a})}\right)^m g(u, u'). \tag{4.71}$$

Then (4.14) becomes

$$F_m(w, z) = \sum_\gamma F_\gamma(w, z) = \sum_\Gamma F_\Gamma(w, z), \tag{4.72}$$

where γ, Γ runs through a residue system mod $1/\mathfrak{d}$ independently. Furthermore, from (4.15) and (4.6) we have the estimate

$$\begin{aligned}&\left|F_\gamma(w, z) - \left(\frac{\mu(\mathfrak{a})}{\varphi(\mathfrak{a})}\right)^m \frac{1}{(2h \log \eta)^m}\frac{1}{(w - 2\pi i(z - \gamma))^m(w' - 2\pi i(z' - \gamma'))^m}\right| \\ &\qquad\qquad < C\frac{1}{\varphi(\mathfrak{a})^m}(VV')^{\frac{1}{2}m},\end{aligned} \tag{4.73}$$

for $\gamma + y$ in $\mathfrak{F}_\gamma$.

5. **Outline of the method.** Our next object is to estimate $D_m(V, V')$. To accomplish this we split $D_m(V, V')$ into four infinite sums and then estimate each sum separately. After introducing (4.72) into (3.3) we obtain

$$
\begin{aligned}
D_m(V, V') &\leqq C \sum_{\gamma} \iint\limits_{\mathfrak{z}'_\gamma} | f^m(w, z) - \sum_{\Gamma} F_\Gamma(w, z) |^2 \, dx_1 \, dx_2 \\
&= C \sum_{\gamma} \iint\limits_{\mathfrak{z}'_\gamma} | f^m(w, z) - F_\gamma(w, z) - \sum_{\Gamma \neq \gamma} F_\Gamma(w, z) |^2 \, dx_1 \, dx_2 \\
&\leqq C \sum_{\gamma} \iint\limits_{\mathfrak{z}'_\gamma} | f^m(w, z) - F_\gamma(w, z) |^2 \, dx_1 \, dx_2 \\
&\qquad + C \sum_{\gamma} \iint\limits_{\mathfrak{z}'_\gamma} | \sum_{\Gamma \neq \gamma} F_\Gamma(w, z) |^2 \, dx_1 \, dx_2 \\
&= S_1 + S_1',
\end{aligned}
\tag{5.11}
$$

say, with γ, Γ mod $1/\mathfrak{d}$ and $N(\mathfrak{a}) \leqq M$. Now let M^* be a positive integer such that $M^* < M$. Also let $\mathfrak{A}$ denote the denominator of $\Gamma\mathfrak{d}$. Then we have

$$
\begin{aligned}
S_1' &= C \sum_{\gamma} \iint\limits_{\mathfrak{z}'_\gamma} | \sum_{\substack{\Gamma \neq \gamma \\ q \leqq M^*}} F_\Gamma(w, z) |^2 \, dx_1 \, dx_2 \\
&\quad + C \sum_{\gamma} \iint\limits_{\mathfrak{z}'_\gamma} | \sum_{\substack{\Gamma \neq \gamma \\ q > M^*}} F_\Gamma(w, z) |^2 \, dx_1 \, dx_2 = C \sum_{\gamma} \iint\limits_{\mathfrak{z}'_\gamma} | \sum_{\substack{\Gamma \neq \gamma \\ q \leqq M^*}} F_\Gamma(w, z) |^2 \, dx_1 \, dx_2 \\
&\quad + C \sum_{\gamma} \iint\limits_{\mathfrak{z}'_\gamma} | \sum_{\substack{\Gamma \\ q > M^*}} F_\Gamma(w, z) - \delta_\gamma F_\gamma(w, z) |^2 \, dx_1 \, dx_2,
\end{aligned}
$$

where γ, Γ are taken mod $1/\mathfrak{d}$, $N(\mathfrak{a}) \leqq M$, $q = N(\mathfrak{A})$, and where $\delta_\gamma = 0$ if $N(\mathfrak{a}) \leqq M^*$ and $\delta_\gamma = 1$ if $M^* < N(\mathfrak{a}) \leqq M$. Hence

$$
\begin{aligned}
S_1' &\leqq C \sum_{\gamma} \iint\limits_{\mathfrak{z}'_\gamma} | \sum_{\substack{\Gamma \neq \gamma \\ q \leqq M^*}} F_\Gamma(w, z) |^2 \, dx_1 \, dx_2 + C \sum_{\gamma} \iint\limits_{\mathfrak{z}'_\gamma} | \sum_{\substack{\Gamma \\ q > M^*}} F_\Gamma(w, z) |^2 \, dx_1 \, dx_2 \\
&\qquad + C \sum_{\gamma} \iint\limits_{\mathfrak{z}'_\gamma} | F_\gamma(w, z) |^2 \, dx_1 \, dx_2 \\
&= S_2 + S_3 + S_4,
\end{aligned}
\tag{5.12}
$$

say, where γ, Γ are taken mod $1/\mathfrak{d}$, $N(\mathfrak{a}) \leqq M$, and in the last summation $M^* < N(\mathfrak{a})$. Combining (5.11) with (5.12) we have finally

$$
D_m(V, V') \leqq S_1 + S_2 + S_3 + S_4 . \tag{5.13}
$$

6. Estimation of the sums.

(i) *Estimation of* S_1. We shall need the following two preliminary estimates:

$$\sum_{\gamma} \iint_{\mathfrak{F}'_\gamma} \left| f\left(\frac{1}{V}, \gamma + y\right) \right|^2 dx_1 \, dx_2 < C(VV')^{1+\epsilon}, \tag{6.11}$$

with γ mod $1/\mathfrak{d}$ and $N(\mathfrak{a}) \leqq M$, and

$$\left| f\left(\frac{1}{V}, \gamma + y\right) \right| < C(VV')^{1+\epsilon}. \tag{6.12}$$

In order to obtain the estimate in (6.11) we first recombine the pieces of the Farey dissection and then employ an argument similar to that used in proving Lemma 5. We get thus

$$\begin{aligned} \sum_{\gamma} \iint_{\mathfrak{F}'_\gamma} \left| f\left(\frac{1}{V}, \gamma + y\right) \right|^2 dx_1 \, dx_2 &= \int_0^1 \int_0^1 \left| f\left(\frac{1}{V}, \gamma + y\right) \right|^2 dx_1 \, dx_2 \\ &= \int_0^1 \int_0^1 \sum_{\omega} \log N(\omega) \exp\left[-(\omega w + \omega' w') + 2\pi i(\omega z + \omega' z')\right] \\ &\quad \cdot \sum_{\omega^*} \log N(\omega^*) \exp\left[-(\omega^* w + \omega^{*\prime} w') - 2\pi i(\omega^* z + \omega^{*\prime} z')\right] dx_1 \, dx_2 \\ &= \sum_{\omega} |\log N(\omega)|^2 \exp(-2\omega w - 2\omega' w') \\ &< C \sum_{\mu} N(\mu)^{\epsilon} \exp(-2\mu w - 2\mu' w'), \end{aligned} \tag{6.131}$$

where γ is taken mod $1/\mathfrak{d}$, $N(\mathfrak{a}) \leqq M$, $\omega \succ 0$, $\omega^* \succ 0$, and $\mu \succ 0$. We now group together those terms of the last sum for which

$$k \leqq \mu < k + 1, \qquad l \leqq \mu' \leqq l + 1.$$

Then, for given k, l, the number of values of μ, μ' is finite and has a maximum value independent of the values of k, l. Hence, for $\mu \succ 0$,

$$\begin{aligned} &\sum_{\mu} N(\mu)^{\epsilon} \exp(-2\mu w - 2\mu' w') \\ &\quad \leqq C + C \sum_{k=1}^{\infty} \sum_{l=1}^{\infty} (kl)^{\epsilon} \exp(-2kw - 2lw') + C \sum_{k=1}^{\infty} k^{\epsilon} e^{-2kw} + C \sum_{l=1}^{\infty} l^{\epsilon} e^{-2lw'} \\ &\quad \leqq C\left(1 + \sum_{k=1}^{\infty} k^{\epsilon} e^{-2kw}\right)\left(1 + \sum_{l=1}^{\infty} l^{\epsilon} e^{-2lw'}\right) \\ &\quad < C(1 + V)^{1+\epsilon}(1 + V')^{1+\epsilon} \\ &\quad < C(VV')^{1+\epsilon}. \end{aligned}$$

The second estimate is derived in a similar manner. We have

$$|f(V^{-1}, \gamma + y)| \leqq C \sum_{\omega} N(\omega)^{\epsilon} \exp(-\omega w - \omega' w') < C(VV')^{1+\epsilon}, \qquad \omega \succ 0.$$

We now return to the sum S_1 defined in (5.11). Introducing the substitutions (3.22), (3.42) and (4.73), we obtain after some simplifications

$$
\begin{aligned}
S_1 \leqq C \sum_\gamma \iint_{\mathfrak{z}'_\gamma} \left|\psi\left(\frac{1}{V}, \gamma + y\right)\right|^2 \left|f\left(\frac{1}{V}, \gamma + y\right)\right|^{2(m-1)} dx_1\, dx_2 \\
+ C(VV')^{2(m-1)} \sum_\gamma \frac{1}{\varphi(\mathfrak{a})^{2(m-1)}} \\
\cdot \iint_{\mathfrak{z}'_\gamma} \frac{\left|\psi\left(\frac{1}{V}, \gamma + y\right)\right|^2}{(1 + 2\pi V\,|\,y\,|)^{2(m-1)}(1 + 2\pi V'\,|\,y'\,|)^{2(m-1)}}\, dx_1\, dx_2 \\
+ C(VV')^m \sum_\gamma \frac{1}{\varphi(\mathfrak{a})^{2m}} \iint_{\mathfrak{z}'_\gamma} dx_1\, dx_2,
\end{aligned} \tag{6.141}
$$

where γ is taken mod $1/\mathfrak{d}$, $N(\mathfrak{a}) \leqq M$, and where we have used the abbreviation

$$
\psi\left(\frac{1}{V}, \gamma + y\right) = f\left(\frac{1}{V}, \gamma + y\right) - \frac{\mu(\mathfrak{a})}{2h \log \eta\, \varphi(\mathfrak{a})} \frac{1}{(w - 2\pi i y)(w' - 2\pi i y')}. \tag{6.142}
$$

The first sum in (6.141) is easily estimated by applying (3.62), (6.11) and (6.12). We get

$$
\begin{aligned}
\sum_\gamma \iint_{\mathfrak{z}'_\gamma} \left|\psi\left(\frac{1}{V}, \gamma + y\right)\right|^2 \left|f\left(\frac{1}{V}, \gamma + y\right)\right|^{2(m-1)} dx_1\, dx_2 \\
\leqq \max_{\text{on all } \mathfrak{z}'_\gamma} \left\{\left|\psi\left(\frac{1}{V}, \gamma + y\right)\right|^2 \left|f\left(\frac{1}{V}, \gamma + y\right)\right|^{2(m-2)}\right\} \\
\cdot \int_0^1 \int_0^1 \left|f\left(\frac{1}{V}, \gamma + y\right)\right|^2 dx_1\, dx_2 \\
\leqq C(VV')^{2\theta + \frac{3}{2} + 4\epsilon}(VV')^{2(m-2)(1+\epsilon)}(VV')^{1+\epsilon} = C(VV')^{2\theta + 2m - \frac{3}{2} + (2m+1)\epsilon},
\end{aligned} \tag{6.15}
$$

with γ mod $1/\mathfrak{d}$ and $N(\mathfrak{a}) \leqq M$.

The estimate of the second sum in (6.141) is more involved. First of all, from (2.3) and (6.142) we get

$$
\begin{aligned}
(VV')^{2(m-1)} \sum_\gamma \frac{1}{\varphi(\mathfrak{a})^{2(m-1)}} \iint_{\mathfrak{z}'_\gamma} \frac{|\,\psi(V^{-1}, \gamma + y)\,|^2}{(1 + 2\pi V\,|\,y\,|)^{2(m-1)}(1 + 2\pi V'\,|\,y'\,|)^{2(m-1)}}\, dx_1\, dx_2 \\
\leqq C(VV')^{2\theta + 2(m-1) + 2\epsilon} \sum_\gamma \frac{N(\mathfrak{a})^{1+4\epsilon}}{\varphi(\mathfrak{a})^{2(m-1)}} \\
\cdot \iint_{\mathfrak{z}'_\gamma} \frac{1}{(1 + 2\pi V\,|\,y\,|)^{2m-3-2\epsilon}(1 + 2\pi V'\,|\,y'\,|)^{2m-3-2\epsilon}}\, dx_1\, dx_2 \\
+ C(VV')^{2m - \frac{3}{2}} \sum_\gamma \frac{N(\mathfrak{a})^{2+2\epsilon}}{\varphi(\mathfrak{a})^{2(m-1)}} \\
\cdot \iint_{\mathfrak{z}'_\gamma} \frac{1}{(1 + 2\pi V\,|\,y\,|)^{2m-3-2\epsilon}(1 + 2\pi V'\,|\,y'\,|)^{2m-3-2\epsilon}}\, dx_1\, dx_2,
\end{aligned} \tag{6.161}
$$

with γ mod $1/\mathfrak{d}$ and $N(\mathfrak{a}) \leqq M$.

Since the estimates of the last two sums are much the same, we shall here treat only the first sum in detail. There are two cases to consider according as $m = 2$ or $m > 2$. In the case $m = 2$ we introduce the change of variables $2\pi V \mid y \mid = U_1$, $2\pi V' \mid y' \mid = U_2$ which carries $\mathfrak{F}'_\gamma$ into $\mathfrak{F}''_\gamma$, say. Now, from (2.52) and (3.41) we have

$$\left| \frac{\partial(U_1, U_2)}{\partial(x_1, x_2)} \right| = \text{absolute value of} \begin{vmatrix} 2\pi V\omega_1 & 2\pi V\omega_2 \\ 2\pi V'\omega_1' & 2\pi V'\omega_2' \end{vmatrix} = \frac{2\pi(VV')}{d^{\frac{1}{2}}}.$$

Hence, using these transformations, we obtain

$$\begin{aligned} &\iint_{\mathfrak{F}'_\gamma} \frac{1}{(1 + 2\pi V \mid y \mid)^{2m-3-2\epsilon}(1 + 2\pi V' \mid y' \mid)^{2m-3-2\epsilon}} \, dx_1 \, dx_2 \\ &\qquad = \frac{d^{\frac{1}{2}}}{2\pi} (VV')^{-1} \iint_{\mathfrak{F}''_\gamma} \frac{1}{(1 + U_1)^{2m-3-2\epsilon}(1 + U_2)^{2m-3-2\epsilon}} \, dU_1 \, dU_2 \\ (6.162) \qquad &\qquad = \frac{d^{\frac{1}{2}}}{2\pi} (VV')^{-1} \iint_{\mathfrak{F}''_\gamma} \frac{1}{(1 + U_1)^{2m-3-2\epsilon}(1 + U_2)^{2m-3-2\epsilon}} \, dU_1 \, dU_2 \\ &\qquad < C(VV')^{-1} \iint_{\mathfrak{F}''_\gamma} dU_1 \, dU_2 = C(VV')^{-1} \iint_{U_2 U_1} dU_1 \, dU_2 \\ &\qquad \leqq C(VV')^{-1} \int_{U_2} \left\{ \int_{U_1} dU_1 \right\} dU_2, \end{aligned}$$

where, because of (3.53), the inner integral in the last member of (6.162) extends over the interval $U_1 \geqq 0$, $1 + U_1 < CM/[N(\mathfrak{a})(1 + U_2)]$. Hence we have

$$(6.163) \qquad \int_{U_1} dU_1 = \begin{cases} 0 & \text{for} \quad \dfrac{CM}{N(\mathfrak{a})(1 + U_2)} \leqq 1, \\[2ex] \dfrac{CM}{N(\mathfrak{a})(1 + U_2)} - 1 & \text{for} \quad \dfrac{CM}{N(\mathfrak{a})(1 + U_2)} \geqq 1. \end{cases}$$

Combining (6.162) with (6.163) now yields

$$\begin{aligned} &\iint_{\mathfrak{F}'_\gamma} \frac{1}{(1 + 2\pi V \mid y \mid)^{2m-3-2\epsilon}(1 + 2\pi V' \mid y' \mid)^{2m-3-2\epsilon}} \, dx_1 \, dx_2 \\ &\qquad < C(VV')^{-1} \int_{1 \leqq 1 + U_2 \leqq p} \left(\frac{CM}{N(\mathfrak{a})(1 + U_2)} - 1 \right) dU_2 \\ (6.164) \qquad &\qquad < C(VV')^{-1} \int_0^p \frac{CM}{N(\mathfrak{a})(1 + U_2)} \, dU_2 = \frac{C(VV')^{-1} M}{N(\mathfrak{a})} \log\left(1 + \frac{CM}{N(\mathfrak{a})}\right) \\ &\qquad < C \frac{(VV')^{-\frac{1}{2}+\epsilon}}{N(\mathfrak{a})}, \end{aligned}$$

where $p = CM/N(\mathfrak{a})$. Therefore

$$(VV')^{2\theta+2(m-1)+2\epsilon} \sum_{\gamma} \frac{N(\mathfrak{a})^{1+4\epsilon}}{\varphi(\mathfrak{a})^{2(m-1)}} \iint\limits_{\mathfrak{F}'_\gamma} \frac{1}{(1+2\pi V|y|)^{2m-3-2\epsilon}(1+2\pi V'|y'|)^{2m-3-2\epsilon}}\, dx_1\, dx_2$$

$$< C(VV')^{2\theta+2m-\frac{1}{2}+3\epsilon} \sum_{\gamma} \frac{N(\mathfrak{a})^{4\epsilon}}{\varphi(\mathfrak{a})^{2(m-1)}}, \tag{6.165}$$

with γ mod $1/\mathfrak{d}$ and $N(\mathfrak{a}) \leqq M$.

It remains to estimate the outer sum in (6.165). For this purpose we observe that to every ideal $\mathfrak{a}$ there correspond $\varphi(\mathfrak{a})$ numbers $\gamma\mathfrak{d}$ that are incongruent mod 1 and such that $\gamma\mathfrak{d}$ has the denominator $\mathfrak{a}$. Hence, there are $\varphi(\mathfrak{a})$ numbers that are incongruent mod $1/\mathfrak{d}$. The sum in question therefore becomes

$$\sum_{\gamma} \frac{N(\mathfrak{a})^{4\epsilon}}{\varphi(\mathfrak{a})^{2(m-1)}} = \sum \frac{N(\mathfrak{a})^{4\epsilon}}{\varphi(\mathfrak{a})^{2m-3}} = \sum \frac{N(\mathfrak{a})^{\epsilon}}{\varphi(\mathfrak{a})} \leqq \sum \frac{\log(N(\mathfrak{a})+1)}{N(\mathfrak{a})^{1-4\epsilon}}$$

$$\leqq \log(M+1) \sum \frac{1}{N(\mathfrak{a})^{1-4\epsilon}} < CM^{\epsilon} M^{4\epsilon} = C(VV')^{\frac{1}{2}\epsilon}, \tag{6.166}$$

with the summation over γ mod $1/\mathfrak{d}$ and $\mathfrak{a}$, where $N(\mathfrak{a}) \leqq M$, and where we have made use of the well-known inequality[25]

$$\varphi(\mathfrak{a}) \geqq C \frac{N(\mathfrak{a})}{\log(N(\mathfrak{a})+1)}. \tag{6.167}$$

From (6.165) and (6.166) we deduce finally that for $m = 2$

$$(VV')^{2\theta+2(m-1)+2\epsilon} \sum_{\gamma} \frac{N(\mathfrak{a})^{1+4\epsilon}}{\varphi(\mathfrak{a})^{2(m-1)}} \iint\limits_{\mathfrak{F}'_\gamma} \frac{1}{(1+2\pi V|y|)^{2m-3-2\epsilon}(1+2\pi V'|y'|)^{2m-3-2\epsilon}}\, dx_1\, dx_2$$

$$< C(VV')^{2\theta+2m-\frac{1}{2}+6\epsilon}, \tag{6.168}$$

with γ mod $1/\mathfrak{d}$ and $N(\mathfrak{a}) \leqq M$. Similarly we may show that the result stated in (6.168) is also valid for $m > 2$. The only essential difference in the proof is that for $m > 2$ we have

$$\iint\limits_{\mathfrak{F}''_\gamma} \frac{1}{(1+U_1)^{2m-3-2\epsilon}(1+U_2)^{2m-3-2\epsilon}}\, dU_1\, dU_2$$

$$< \iint\limits_{-\infty}^{\infty} \frac{1}{(1+U_1)^{2m-3-2\epsilon}(1+U_2)^{2m-3-2\epsilon}}\, dU_1 dU_2 = C.$$

The estimate of the second sum in the right side of (6.161) is also obtained in a similar manner. We get

$$(VV')^{2m-\frac{1}{2}} \sum_{\gamma} \frac{N(\mathfrak{a})^{2+2\epsilon}}{\varphi(\mathfrak{a})^{2(m-1)}} \iint\limits_{\mathfrak{F}'_\gamma} \frac{1}{(1+2\pi V|y|)^{2m-3-2\epsilon}(1+2\pi V'|y'|)^{2m-3-2\epsilon}}\, dx_1\, dx_2$$

$$< C(VV')^{2m-\frac{1}{2}+4\epsilon}, \tag{6.17}$$

[25] Rdm. I, p. 147.

with γ mod $1/\mathfrak{d}$ and $N(\mathfrak{a}) \leqq M$. We now consider the last sum in (6.141). A rough estimate will suffice.

$$\begin{aligned}(VV')^m \sum_{\gamma} \frac{1}{\varphi(\mathfrak{a})^{2m}} \iint_{\mathfrak{F}'_\gamma} dx_1\,dx_2 &< (VV')^m \sum_{\gamma} \frac{1}{\varphi(\mathfrak{a})^{2m}} \int_0^1 \int_0^1 dx_1\,dx_2 \\ &= (VV')^m \sum_{\gamma} \frac{1}{\varphi(\mathfrak{a})^{2m}} = (VV')^m \sum \frac{1}{\varphi(\mathfrak{a})^{2m-1}} \\ &\leqq C(VV')^m \sum \frac{1}{\varphi(\mathfrak{a})^3} < C(VV')^m,\end{aligned} \tag{6.18}$$

with the summation over γ mod $1/\mathfrak{d}$ and $\mathfrak{a}$, where $N(\mathfrak{a}) \leqq M$.

Finally we combine the results of (6.141), (6.15), (6.161), (6.168), (6.17), (6.18), (6.19). Since $\frac{1}{2} \leqq \theta \leqq 1$, we may verify immediately that

$$\begin{aligned}S_1 \leqq C(VV')^{2\theta+2m-\frac{5}{2}+(2m+1)\epsilon} + C(VV')^{2m-\frac{3}{2}+4\epsilon} + C(VV')^m \\ \leqq C(VV')^{2\theta+2m-\frac{5}{2}+\epsilon},\end{aligned} \tag{6.19}$$

where we have replaced $(2m+1)\epsilon$ by ϵ.

(ii) *Estimation of* S_2.[26] From the estimation of $F_\gamma(w, z)$ in (4.73) we have after an elementary modification

$$\begin{aligned}F_\Gamma(w, z) = \left(\frac{\mu(\mathfrak{A})}{\varphi(\mathfrak{A})}\right)^m \frac{1}{(2h \log \eta)^m} \frac{1}{(w - 2\pi i(z-\Gamma))^m (w' - 2\pi i(z' - \Gamma'))^m} \\ + B \frac{1}{\varphi(\mathfrak{A})^m} (VV')^{\frac{1}{2}m},\end{aligned} \tag{6.21}$$

where B is an absolute constant such that $|B| \leqq C$. Before introducing (6.21) into S_2 (defined in (5.12)) we apply the Cauchy-Schwarz inequality and then interchange the signs of summation. We get thus

$$\begin{aligned}S_2 &\leqq CM^{*2} \sum_{\gamma} \iint_{\mathfrak{F}'_\gamma} \sum_{\Gamma \neq \gamma} |F_\Gamma(w, z)|^2\,dx_1\,dx_2 \\ &= CM^{*2} \sum_{\substack{\Gamma \\ \Gamma \neq \gamma}} \sum_{\gamma} \iint_{\mathfrak{F}'_\gamma} |F_\Gamma(w, z)|^2\,dx_1\,dx_2 \\ &= CM^{*2} \sum_{\Gamma} \iint_{\mathfrak{G}'_\Gamma} |F_\Gamma(w, z)|^2\,dx_1\,dx_2 \\ &< CM^{*2} \sum_{\Gamma} \frac{1}{\varphi(\mathfrak{A})^{2m}} \iint_{\mathfrak{G}'_\Gamma} \frac{1}{(w + 2\pi|z - \Gamma|)^{2m}(w' + 2\pi|z' - \Gamma'|)^{2m}}\,dx_1\,dx_2 \\ &\quad + CM^{*2}(VV')^m \sum_{\Gamma} \frac{1}{\varphi(\mathfrak{A})^{2m}} \iint_{\mathfrak{G}'_\Gamma} dx_1\,dx_2 = S_2' + S_2'',\end{aligned} \tag{6.22}$$

[26] The idea of the method employed in this paragraph was kindly suggested to the author by Professor Rademacher.

say, with γ, Γ mod $1/\mathfrak{d}$, $N(\mathfrak{a}) \leqq M$, and $N(\mathfrak{A}) \leqq M^*$ and $\mathfrak{G}'_\Gamma$ denoting the area that remains after $\mathfrak{F}'_\Gamma$ is removed from the fundamental parallelogram.

We first estimate S'_2. In view of (2.52) we introduce the substitution

$$Y = z - \Gamma = \omega_1 x_1 + \omega_2 x_2 - \Gamma, \qquad Y' = z' - \Gamma' = \omega'_1 x_1 + \omega'_2 x_2 - \Gamma'$$

and get

$$\iint\limits_{\mathfrak{G}'_\Gamma} \frac{1}{(w + 2\pi \mid z - \Gamma \mid)^{2m}(w' + 2\pi \mid z' - \Gamma' \mid)^{2m}}\, dx_1\, dx_2$$

$$\text{(6.23)} \qquad < (VV')^{2m} \iint\limits_{\mathfrak{G}'_\Gamma} \frac{1}{(1 + 2\pi \mid Y \mid w^{-1})^{2m}(1 + 2\pi \mid Y' \mid w'^{-1})^{2m}}\, dx_1\, dx_2$$

$$= \frac{d^{\frac{1}{2}}}{2\pi} (VV')^{2m-1} \iint\limits_{\mathfrak{G}''_\Gamma} \frac{1}{(1 + U_1)^{2m}(1 + U_2)^{2m}}\, dU_1\, dU_2,$$

where $\mathfrak{G}''_\Gamma$ denotes the area that remains after $\mathfrak{F}'_\Gamma$ is removed from the plane.

Now by Lemma 1 we have for $\Gamma + Y$ not in $\mathfrak{F}'_\Gamma$ and hence for $\Gamma + Y$ in $\mathfrak{G}''_\Gamma$

$$(1 + M \mid Y \mid)(1 + M \mid Y' \mid) > C \frac{M}{N(\mathfrak{A})}.$$

Therefore, by (3.22), (3.42) and (3.52) we have for $\Gamma + Y$ in $\mathfrak{G}''_\Gamma$

$$(1 + U_1)(1 + U_2) > C \frac{M}{N(\mathfrak{A})} = P, \qquad U_1 \geqq 0,\ U_2 \geqq 0,$$

where the letter P has been introduced for the sake of brevity.

The convergence of the integral in (6.23) is assured since $2m \geqq 4$. Hence we may write

$$\iint\limits_{\mathfrak{G}''_\Gamma} < \int_0^\infty \frac{1}{(1+U_1)^{2m}} \left\{ \int_{U_1} \frac{dU_1}{(1 + U_1)^{2m}} \right\} dU_2,$$

where the inner integral extends over the interval $U_1 \geqq 0$, $1 + U_1 > P/(1 + U_2)$. An elementary integration now yields

$$\int_{U_1} \frac{dU_1}{(1 + U_1)^{2m}} = \begin{cases} \dfrac{1}{2m - 1} & \text{for } \dfrac{P}{1 + U_2} < 1, \\[2ex] \dfrac{1}{2m - 1} \left(\dfrac{1 + U_2}{P} \right)^{2m-1} & \text{for } \dfrac{P}{1 + U_2} \geqq 1. \end{cases}$$

Hence

$$\text{(6.24)} \qquad \iint\limits_{\mathfrak{G}''_\Gamma} < C \int_{1 \leqq 1+U_2 > P} \frac{dU_2}{(1 + U_2)^{2m}} + C \int_{1 \leqq 1+U_2 \leqq P} \frac{1}{P^{2m-1}} \frac{1}{1 + U_2}\, dU_2$$

$$< \frac{C}{P^{2m-1}} + \frac{C}{P^{2m-1}} \log (P + 1) < C(VV')^{-m+\frac{1}{2}+\epsilon} N(\mathfrak{A})^{2m-1}.$$

Introducing (6.23) and (6.24) into S_2' gives

$$(6.25)\quad \begin{aligned} S_2' &< CM^{*2}(VV')^{m-\frac{1}{2}+\epsilon}\sum_{\Gamma}\frac{N(\mathfrak{A})^{2m-1}}{\varphi(\mathfrak{A})^{2m}} < CM^{*2}(VV')^{m-\frac{1}{2}+\epsilon}\sum\left(\frac{N(\mathfrak{A})}{\varphi(\mathfrak{A})}\right)^{2m-1} \\ &< CM^{*2}(VV')^{m-\frac{1}{2}+\epsilon}(\log(M^*+1))^{2m-1}\sum 1 < CM^{*3+\epsilon}(VV')^{m-\frac{1}{2}+\epsilon}, \end{aligned}$$

where the first summation is over Γ mod $1/\mathfrak{d}$ and the others over $\mathfrak{A}$, for $N(\mathfrak{A}) \leqq M^*$.

The estimation of S_2'' involves no new difficulty. We have immediately

$$(6.26)\quad \begin{aligned} S_2'' &= M^{*2}(VV')^m \sum_{\Gamma}\frac{1}{\varphi(\mathfrak{A})^{2m}}\iint_{\mathfrak{G}_\Gamma'} dx_1\,dx_2 \\ &< M^{*2}(VV')^m\sum_{\Gamma}\frac{1}{\varphi(\mathfrak{A})^{2m}}\int_0^1\int_0^1 dx_1\,dx_2 = M^{*2}(VV')^m\sum_{\Gamma}\frac{1}{\varphi(\mathfrak{A})^{2m}} \\ &< C(VV')^m M^{*2+\epsilon}, \end{aligned}$$

with Γ mod $1/\mathfrak{d}$ and $N(\mathfrak{A}) \leqq M^*$.

(iii) *Estimation of* S_3. In order to estimate S_3 we recombine the pieces of the Farey dissection and obtain

$$S_3 = \int_0^1\int_0^1 \Big|\sum_{\Gamma} F_\Gamma(w, z)\Big|^2 dx_1\,dx_2,$$

where Γ is taken mod $1/\mathfrak{d}$ and $N(\mathfrak{A}) > M^*$, with

$$(6.31)\quad \begin{aligned} \sum_{\Gamma} F_\Gamma(w, z) &= \frac{d^{\frac{1}{2}}}{(2h\log\eta)^m\Gamma^2(m)}\sum_{\mu} N(\mu)^{m-1} \\ &\quad\cdot\sum A_m(\mu;\mathfrak{A})\exp[-\mu w - \mu' w' + 2\pi i(\mu z + \mu' z')], \end{aligned}$$

where the summation is over Γ mod $1/\mathfrak{d}$, $\mu > 0$, and $\mathfrak{A}$ for $N(\mathfrak{A}) > M^*$, and where we have used (4.11).

The estimation of $A_m(\mu;\mathfrak{A})$ depends on (4.12) and (4.13). We first observe that $A_m(\mu;\mathfrak{A})$ is zero if $\mathfrak{A}$ is not square-free. If $\mathfrak{A}$ is square-free we put $\mathfrak{A} = \mathfrak{A}_1\mathfrak{A}_2$, where $\mathfrak{A}_1 \mid \mu$ and $(\mathfrak{A}_2, \mu) = 1$. Then, making use of (6.167), we have

$$|A_m(\mu;\mathfrak{A})| = \frac{1}{\varphi(\mathfrak{A})^m}\left|\sum N(\mathfrak{b})\mu\left(\frac{\mathfrak{A}_1}{\mathfrak{b}}\right)\right| = \frac{\varphi(\mathfrak{A}_1)}{\varphi(\mathfrak{A})^m} \leqq C\,\frac{N(\mathfrak{A})^\epsilon}{N(\mathfrak{A}_1)^{m-1}N(\mathfrak{A}_2)^m},$$

where the summation is over $\mathfrak{b} \mid \mathfrak{A}_1$. Hence

$$\begin{aligned} \Big|\sum_{\substack{\mathfrak{A} \\ N(\mathfrak{A})>M^*}} A_m(\mu;\mathfrak{A})\Big| &\leqq C\sum_{\substack{N(\mathfrak{A}_1\mathfrak{A}_2)>M^* \\ \mathfrak{A}_1\mathfrak{A}_2\text{ square free} \\ \mathfrak{A}_1\mid\mu}}\frac{N(\mathfrak{A}_1)^\epsilon N(\mathfrak{A}_2)^\epsilon}{N(\mathfrak{A}_1)^{m-1}N(\mathfrak{A}_2)^m} \\ &\leqq C\sum_{\substack{\mathfrak{A}_1\mid\mu \\ \mathfrak{A}_1\text{ square free}}}\frac{1}{N(\mathfrak{A}_1)^{m-1-\epsilon}}\sum_{N(\mathfrak{A}_2)>\frac{M^*}{N(\mathfrak{A}_1)}}\frac{1}{N(\mathfrak{A}_2)^{m-\epsilon}} \end{aligned}$$

$$
\begin{aligned}
&< C \sum_{\substack{\mathfrak{A}_1 \mid \mu \\ N(\mathfrak{A}_1) > M^*}} \frac{1}{N(\mathfrak{A}_1)^{m-1-\epsilon}} \\
&\qquad + C \sum_{\substack{\mathfrak{A}_1 \mid \mu \\ \mathfrak{A}_1 \text{ square free}}} \frac{1}{N(\mathfrak{A}_1)^{m-1-\epsilon}} \left(\frac{M^*}{N(\mathfrak{A}_1)} \right)^{-m+1+\epsilon} \\
&< CM^{*^{-m+1+\epsilon}} \sum_{\substack{\mathfrak{A}_1 \mid \mu \\ \mathfrak{A}_1 \text{ square free}}} 1 = CM^{*^{-m+1+\epsilon}} \sum_{\mathfrak{b} \mid \mu} \mu(\mathfrak{b})^2 \\
&= CM^{*^{-m+1+\epsilon}} \prod_{\mathfrak{P} \mid \mu} (1 + \mu(\mathfrak{P}))^2 = CM^{*^{-m+1+\epsilon}} 2^t \\
&\qquad < CN(\mu)^{\epsilon} M^{*^{-m+1+\epsilon}},
\end{aligned} \tag{6.32}
$$

where t denotes the number of different prime ideal factors of μ.[27]

Introducing (6.31) and (6.32) into the integrand of S_3 we get

$$
\begin{aligned}
S_3 &\leqq CM^{*^{-2m+2+2\epsilon}} \\
&\cdot \int_0^1 \int_0^1 \left| \sum_{\mu} N(\mu)^{m-1+\epsilon} \exp\left[-(\mu w + \mu' w') + 2\pi i(\mu z + \mu' z') \right] \right|^2 dx_1\, dx_2, \quad \mu \succ 0.
\end{aligned}
$$

Lemma 5 is now clearly applicable. We obtain

$$
\begin{aligned}
S_3 &\leqq CM^{*^{-2m+2+2\epsilon}} \sum_{\mu} N(\mu)^{2m-2+2\epsilon} \exp(-2\mu w - 2\mu' w') \\
&\leqq CM^{*^{-2m+2+2\epsilon}} \max \{ N(\mu)^{2m-2+2\epsilon} \exp(-\mu w - \mu' w') \} \\
&\qquad\qquad \cdot \sum_{\mu} \exp(-\mu w - \mu' w') \\
&= CM^{*^{-2m+2+2\epsilon}} (ww')^{-2m+2-2\epsilon} \sum_{\mu} \exp(-\mu w - \mu' w') \\
&\leqq CM^{*^{-2m+2+2\epsilon}} (VV')^{2m-2+2\epsilon} (VV')^{1+\epsilon} \\
&= C(VV')^{2m-1+\epsilon} M^{*^{-2m+2+2\epsilon}},
\end{aligned} \tag{6.33}
$$

where $\mu \succ 0$ and where we have calculated $\max \{N(\mu)^{2m-2+2\epsilon} \exp(-\mu w - \mu' w')\}$ by means of an easy differentiation.

(iv) *Estimation of* S_4. By (4.73) and (5.12) we have

$$
\begin{aligned}
S_4 &\leqq C \sum_{\gamma} \frac{1}{\varphi(\mathfrak{a})^{2m}} \iint_{\mathfrak{B}'_\gamma} \frac{1}{(w + 2\pi \mid y \mid)^{2m} (w' + 2\pi \mid y' \mid)^{2m}}\, dx_1\, dx_2 \\
&\qquad + C(VV')^m \sum_{\gamma} \frac{1}{\varphi(\mathfrak{a})^{2m}} \iint_{\mathfrak{B}'_\gamma} dx_1\, dx_2
\end{aligned} \tag{6.41}
$$

[27] For a proof of the estimate $2^t < CN(\mu)^{\epsilon}$ see Landau, loc. cit., pp. 188–189, Theorem 221.

with γ mod $1/\mathfrak{d}$ and $M^* < N(\mathfrak{a}) \leqq M$. The two sums in (6.41) may be estimated by using methods already explained in detail. Thus we find, putting $p = N(\mathfrak{a})$, that

$$\sum_{\substack{\gamma \\ M^* < p \leqq M}} \frac{1}{\varphi(\mathfrak{a})^{2m}} \iint_{\mathfrak{z}'_\gamma} \frac{1}{(w + 2\pi \mid y \mid)^{2m}(w' + 2\pi \mid y' \mid)^{2m}}\, dx_1\, dx_2$$

$$= (VV')^{2m} \sum_{\substack{\gamma \\ M^* < p \leqq M}} \frac{1}{\varphi(\mathfrak{a})^{2m}} \iint_{\mathfrak{z}'_\gamma} \frac{1}{(1 + 2\pi V \mid y \mid)^{2m}(1 + 2\pi V' \mid y' \mid)^{2m}}\, dx_1\, dx_2$$

$$= \frac{d^{\frac{1}{2}}}{2\pi} (VV')^{2m-1} \sum_{M^* < p \leqq M} \frac{1}{\varphi(\mathfrak{a})^{2m-1}} \iint_{\mathfrak{z}''_\gamma} \frac{1}{(1 + U_1)^{2m}(1 + U_2)^{2m}}\, dU_1\, dU_2 \tag{6.42}$$

$$< C(VV')^{2m-1} \sum_{p > M^*} \frac{1}{\varphi(\mathfrak{a})^{2m-1}} \int_{-\infty}^{\infty}\!\!\int \frac{1}{(1 + U_1)^{2m}(1 + U_2)^{2m}}\, dU_1\, dU_2$$

$$< C(VV')^{2m-1} \sum_{p > M^*} \frac{1}{N(\mathfrak{a})^{2m-1-(2m-1)\epsilon}}$$

$$< C(VV')^{2m-1} M^{*-2m+(2m-1)\epsilon}.$$

For the second sum in (6.41) we obtain

$$(VV')^m \sum_{\substack{\gamma \\ M^* < p \leqq M}} \frac{1}{\varphi(\mathfrak{a})^{2m}} \iint_{\mathfrak{z}'_\gamma} dx_1\, dx_2$$

$$\begin{aligned} &< (VV')^m \sum_{\substack{\gamma \\ M^* < p \leqq M}} \frac{1}{\varphi(\mathfrak{a})^{2m}} \int_0^1 \int_0^1 dx_1\, dx_2 \\ &< C(VV')^m \sum_{M^* < p \leqq M} \frac{1}{\varphi(\mathfrak{a})^{2m-1}} \\ &< C(VV')^m \sum_{p > M^*} \frac{1}{N(\mathfrak{a})^{2m-1-(2m-1)\epsilon}} \\ &< C(VV')^m M^{*-2m+(2m-1)\epsilon}. \end{aligned} \tag{6.43}$$

7. **The main theorem.** We proceed to the estimate of $D_m(V, V')$. Combining the results of (5.11), (5.12), (5.13), (6.19), (6.22), (6.25), (6.26), (6.33), (6.41), (6.42), (6.43), we have

$$\begin{aligned} D_m(V, V') \leqq {} & C(VV')^{2\theta+2m-\frac{1}{2}+\epsilon} + CM^{*3+\epsilon}(VV')^{m-\frac{1}{2}+\epsilon} \\ & + C(VV')^m M^{*2+\epsilon} + C(VV')^{2m-1+\epsilon} M^{*-2m+2+2\epsilon} \\ & + C(VV')^{2m-1} M^{*-2m+\epsilon} + C(VV')^m M^{*-2m+(2m-1)\epsilon}. \end{aligned} \tag{7.11}$$

Let us take $M^* = (VV')^{\frac{1}{4}}$. Then (using the fact that $\frac{1}{2} \leqq \theta \leqq 1$) we may easily verify that the right member of (7.1) is dominated by the first term. Hence

$$D_m(V, V') \leqq C(VV')^{2\theta+2m-\frac{3}{2}+\epsilon}. \tag{7.12}$$

We recall that up to the present we had assumed that $\eta_1^{-1} \leqq V'V^{-1} \leqq \eta_1$ (see §3). We shall now remove this restriction by showing that (7.12) is true in general. For this purpose we put

$$V_0 = V\eta_1^{-k}, \qquad V_0' = V'\eta_1^{-k}, \tag{7.21}$$

$$\eta_1^{-1} \leqq \frac{V_0'}{V_0} \leqq \eta_1, \tag{7.22}$$

where V and V' now denote two arbitrary real numbers and

$$k = -\left[\frac{\log V' - \log V}{2\log \eta_1} + \frac{1}{2}\right].$$

It follows from (7.22) that

$$D_m(V_0, V_0') \leqq C(V_0V_0')^{2\theta+2m-\frac{3}{2}+\epsilon}$$

in the same way that (7.12) follows from (3.21). Furthermore from (7.21) we have $V_0V_0' = VV'$. Therefore, in order to finish the proof it is only necessary to show that $D_m(V_0, V_0') = D_m(V, V')$. This can be seen directly from the definition (3.11). For, all totally positive integers μ which satisfy the inequalities $0 < \mu \leqq V_0$, $0 < \mu' \leqq V_0'$ also satisfy the inequalities $0 < \mu\eta_1^k \leqq V$, $0 < \mu'\eta_1^{-k} \leqq V'$, and conversely. If we now write μ in place of $\mu\eta_1^k$, the proof is complete.

We collect our results in

THEOREM 1. *Let θ denote the upper bound of the real parts of all the zeros of all the Hecke $\zeta(s, \lambda)$-functions in a real quadratic field $k(d^{\frac{1}{2}})$. Then for $m \geqq 2$ we have*

$$\sum_\mu \left(A_m(\mu) - \frac{d^{\frac{1}{2}}}{(2h\log\eta)^m\Gamma^2(m)} N(\mu)^{m-1}\mathfrak{S}_m(\mu)\right)^2 \leqq C(VV')^{2\theta+2m-\frac{3}{2}+\epsilon}, \tag{7.3}$$

where $0 < \mu \leqq V$, $0 < \mu' \leqq V'$, and where $A_m(\mu)$ and $\mathfrak{S}_m(\mu)$ are defined by (2.26) *and* (2.42), *respectively. The letter h denotes the class number and the letter η denotes the fundamental unit of the field.*

Our next object is to interpret this theorem so as to yield the additive prime number theorem mentioned in the introduction. For this purpose it is necessary to study $\mathfrak{S}_2(\mu)$. From (2.43) we obtain

$$\mathfrak{S}_2(\mu) = \prod_{\mathfrak{P}\nmid\mu}\left(1 - \frac{1}{(N(\mathfrak{P})-1)^2}\right)\prod_{\mathfrak{P}\mid\mu}\left(1 + \frac{1}{N(\mathfrak{P})-1}\right). \tag{7.41}$$

Since the infinite product in (7.31) converges, $\mathfrak{S}_2(\mu)$ vanishes if and only if a factor vanishes. Clearly this occurs if and only if there exists a prime ideal $\mathfrak{P}^*$ such that $N(\mathfrak{P}^*) = 2$ and $\mathfrak{P}^* \nmid \mu$. For a clearer statement of this fact we find it convenient to adopt the following terminology. Let $\mathfrak{l}$ denote the product of all the different prime ideals $\mathfrak{P}$ such that $N(\mathfrak{P}) = 2$. If such prime ideals do not exist, let $\mathfrak{l}$ denote the principal ideal (1). An integer μ is said to be "even" if $\mathfrak{l} \mid \mu$; it is said to be "odd" if $(\mathfrak{l}, \mu) = 1$. In other words, if μ is "even" the prime factors with norm 2 (provided any exist) all divide μ; but, if μ is "odd" no prime factor with norm 2 divides μ. Using this terminology we may now state that $\mathfrak{S}_2(\mu)$ vanishes if and only if μ is "not-even". If μ is "even" $\mathfrak{S}_2(\mu)$ does not vanish and we pick out from (7.31) those factors which arise from prime ideals with norm 2. The product of these factors is 2^g, where g is the number of different prime factors with the norm 2. It is well known that $g = 1 + (d \mid 2)$, where $(d \mid 2)$ denotes the Legendre symbol.[28] Therefore (7.31) becomes

$$\begin{aligned} \mathfrak{S}_2(\mu) &= 2 \cdot 2^{(d|2)} \prod_{\mathfrak{P}} \left(1 - \frac{1}{(N(\mathfrak{P}) - 1)^2}\right) \prod_{\mathfrak{P}|\mu} \frac{1 + [N(\mathfrak{P}) - 1]^{-1}}{1 - [N(\mathfrak{P}) - 1]^{-2}} \\ &> C \prod_{\mathfrak{P}|\mu} \frac{N(\mathfrak{P}) - 1}{N(\mathfrak{P}) - 2}, \end{aligned} \tag{7.42}$$

where $N(\mathfrak{P}) > 2$.

We are now in the position to derive our prime number theorem. We shall call a totally positive "even" integer μ a "non-Goldbach" number if there does not exist a pair of totally positive primes ω_1, ω_2 such that $\mu = \omega_1 + \omega_2$. By (2.26) and (7.32) we have for a non-Goldbach number μ

$$\begin{aligned} &\left(A_2(\mu) - \frac{d^{\frac{1}{2}}}{(2h \log \eta)^2 \Gamma^2(2)} N(\mu) \mathfrak{S}_2(\mu)\right)^2 \\ &\qquad = C^2 N(\mu)^2 \prod_{\mathfrak{P}|\mu} \left(\frac{N(\mathfrak{P}) - 1}{N(\mathfrak{P}) - 2}\right)^2 > C N(\mu)^2, \end{aligned} \tag{7.51}$$

where $N(\mathfrak{P}) > 2$. Let $N(a, b)$ be the number of non-Goldbach numbers μ such that $a < \mu \leqq 2a$, $b < \mu' \leqq 2b$. Then (7.41) clearly implies

$$\sum_{\mu} \left(A_2(\mu) - \frac{d^{\frac{1}{2}}}{(2h \log \eta)^2 \Gamma^2(2)} N(\mu) \mathfrak{S}_2(\mu)\right)^2 \geqq C N(\tfrac{1}{2}V, \tfrac{1}{2}V')(\tfrac{1}{4}VV')^2, \tag{7.52}$$

where $0 < \mu \leqq V$, $0 < \mu' \leqq V'$. Therefore, by (7.3) and (7.52),

$$N(\tfrac{1}{2}V, \tfrac{1}{2}V') \leqq C(VV')^{2\theta - \frac{1}{2} + 2\epsilon}.$$

In order to estimate the number $P(V, V')$ of non-Goldbach numbers μ satisfying the inequalities $0 < \mu \leqq V$, $0 < \mu' \leqq V'$ we subdivide the rectangle with vertices at $(0, 0)$, $(V, 0)$, $(0, V')$, (V, V') into an infinite number of smaller rectangles by drawing lines through the points $(\frac{1}{2}V, 0)$, $(\frac{1}{4}V, 0)$, $(\frac{1}{8}V, 0)$, $\cdots$

[28] See Hecke, loc. cit., p. 110 and pp. 186–187.

parallel to one of the sides and through the points $(0, \frac{1}{2}V')$, $(0, \frac{1}{4}V')$, $(0, \frac{1}{8}V')$, $\cdots$ parallel to the other side. We then have

$$\begin{aligned} P(V, V') &= \sum_{m=1}^{\infty} \sum_{n=1}^{\infty} N\left(\frac{V}{2^m}, \frac{V}{2^n}\right) \\ &\leqq C(VV')^{2\theta-\frac{1}{2}+\epsilon} \sum_{m=1}^{\infty} \sum_{n=1}^{\infty} \frac{1}{(2^m)^{2\theta-\frac{1}{2}+2\epsilon}} \frac{1}{(2^n)^{2\theta-\frac{1}{2}+\epsilon}} \\ &\leqq C(VV')^{2\theta-\frac{1}{2}+\epsilon} \sum_{m=1}^{\infty} \sum_{n=1}^{\infty} \frac{1}{2^{\frac{1}{2}m} 2^{\frac{1}{2}n}} \\ &\leqq C(VV')^{2\theta-\frac{1}{2}+\epsilon}. \end{aligned}$$

This completes the proof of

THEOREM 2. *Let θ denote the upper bound of the real parts of all the zeros of all the Hecke $\zeta(s, \lambda)$-functions in a real quadratic field. If*

$$\theta < \tfrac{3}{4},$$

then almost all totally positive even integers in the field can be represented as the sum of two totally positive primes. More precisely, if $P(V, V')$ is the number of totally positive even integers μ in the rectangle $0 < \mu \leqq V$, $0 < \mu' \leqq V'$ which cannot be represented as the sum of two totally positive primes, then

$$\lim_{\substack{V \to \infty \\ V' \to \infty}} \frac{P(V, V')}{(VV')^{2\theta-\frac{1}{2}+\epsilon}} = 0.$$

UNIVERSITY OF PENNSYLVANIA.

A THEORY OF ANALYTIC FUNCTIONS IN LINEAR ASSOCIATIVE ALGEBRAS

By James A. Ward

1. **Introduction.** In 1893, Scheffers [50][1] published a pioneering paper in the theory of functions in linear algebras. He showed that, for a distributive algebra with a unit element over a complex field, a necessary and sufficient condition that the derivative in the ordinary sense exist is that the algebra be commutative, and for the integral to exist it is necessary and sufficient that the algebra also be associative. Many attempts have been made to define derivative, integral and analytic function in such a way that at least part of the ordinary function theory might carry over. Some of the definitions are very ingenious.

In the present paper, derivative is defined for every function, and a function is called analytic if its derivative lies in a certain algebra $\mathfrak{D} \supseteq \mathfrak{A}$, which becomes $\mathfrak{A}$ when $\mathfrak{A}$ is commutative. By thus allowing the derivative to be outside the original algebra $\mathfrak{A}$, Scheffers' limitations are avoided. In fact, this general definition of analytic function is shown to be equivalent to that of Hausdorff [26], who did not succeed in obtaining a definition of derivative to go with it.

The author has concluded the paper with as complete a bibliography on analytic functions of linear algebras as he has been able to obtain.

The present paper is taken from the author's doctor's thesis at the University of Wisconsin, written under the direction of Professor C. C. MacDuffee, whom the author wishes to thank for his interest and help.

2. **The derived algebra.** Let $\mathfrak{A}$ be a finite linear associative algebra over a field $\mathfrak{F}$. Let $\epsilon_1, \epsilon_2, \cdots, \epsilon_n$ be an $\mathfrak{F}$-basis, where ϵ_1 is a unit element. Let

$$\epsilon_i \epsilon_j = \sum_k c_{ijk} \epsilon_k, \qquad c_{ijk} \in \mathfrak{F}.$$

Denote by R_i the matrix (c_{isr}), and by S_i the matrix (c_{ris}), where r is the row-index and s the column-index. If

$$\alpha = a_1\epsilon_1 + a_2\epsilon_2 + \cdots + a_n\epsilon_n, \qquad a_i \in \mathfrak{F},$$

is any number of $\mathfrak{A}$, then the correspondences

$$\alpha \leftrightarrow R(\alpha) = a_1R_1 + a_2R_2 + \cdots + a_nR_n,$$

$$\alpha \leftrightarrow S(\alpha) = a_1S_1 + a_2S_2 + \cdots + a_nS_n$$

Received March 26, 1940; presented to the American Mathematical Society, April 14, 1939.

[1] Numbers in brackets refer to papers in the bibliography.

are isomorphisms under both addition and multiplication, the well-known first and second regular representations of $\mathfrak{A}$ by matrices.

Let us now consider a new algebra $\bar{\mathfrak{A}}$ over $\mathfrak{F}$ with the basis numbers $\bar{\epsilon}_1, \bar{\epsilon}_2, \cdots, \bar{\epsilon}_n$ and the rule of multiplication

$$\bar{\epsilon}_i\bar{\epsilon}_j = \sum \bar{c}_{ijk}\bar{\epsilon}_k, \qquad \bar{c}_{ijk} = c_{jik} \in \mathfrak{F}.$$

If we set

$$\bar{\alpha} = a_1\bar{\epsilon}_1 + a_2\bar{\epsilon}_2 + \cdots + a_n\bar{\epsilon}_n,$$

the correspondence $\alpha \leftrightarrow \bar{\alpha}$ is an anti-isomorphism, or reciprocal isomorphism. That is, $\alpha \leftrightarrow \bar{\alpha}$ and $\beta \leftrightarrow \bar{\beta}$ imply

$$\alpha + \beta \leftrightarrow \bar{\alpha} + \bar{\beta}, \qquad \alpha\beta \leftrightarrow \bar{\beta}\bar{\alpha}.$$

The algebra $\bar{\mathfrak{A}}$ is called the reciprocal algebra of the algebra $\mathfrak{A}$.

Let us consider the regular representations of $\bar{\mathfrak{A}}$. Denote by $\bar{R}_i$ the matrix $(\bar{c}_{isr})$, and by $\bar{S}_i$ the matrix $(\bar{c}_{ris})$. But $\bar{R}_i = (c_{sir}) = S_i^T$, and $\bar{S}_i = (c_{irs}) = R_i^T$. That is, the matrices of the first regular representation of $\bar{\mathfrak{A}}$ are the transposes of the matrices of the second regular representation of $\mathfrak{A}$, and vice versa.

If $\mathfrak{A}$ is commutative, and only then, $\bar{\mathfrak{A}}$ coincides with $\mathfrak{A}$, $\bar{R}_i$ with R_i, and $\bar{S}_i$ with S_i, so that $R_i^T = S_i$.

Since $\mathfrak{A}$ contains a unit element, it is clear that both $\mathfrak{A}$ and $\bar{\mathfrak{A}}$ are (isomorphic with) subalgebras of the direct product $\mathfrak{A} \times \bar{\mathfrak{A}}$. The algebra $\mathfrak{D}$ of lowest order in which both $\mathfrak{A}$ and $\bar{\mathfrak{A}}$ can be imbedded will be called the first derived algebra[2] of $\mathfrak{A}$.

This algebra can be explicitly represented by employing the regular representations of $\mathfrak{A}$ and $\bar{\mathfrak{A}}$. Consider the linear set

$$\sum a_{ij}R_i\bar{R}_j, \qquad a_{ij} \in \mathfrak{F}, \tag{1}$$

of $n \times n$ matrices. Since $\bar{R}_j = S_j^T$, and for every i and j it is true that[3]

$$R_iS_j^T = S_j^TR_i, \tag{2}$$

we have

$$R_i\bar{R}_jR_k\bar{R}_l = R_iR_k\bar{R}_j\bar{R}_l,$$

so that the linear set (1) is closed under multiplication, and hence constitutes an algebra $\mathfrak{D}$ over $\mathfrak{F}$. We shall denote by

$$D_1, D_2, \cdots, D_p$$

[2] This algebra was considered by Jacobson for a more general $\mathfrak{A}$, and called by him the "enveloping algebra of the left and right multiplications of $\mathfrak{A}$". N. Jacobson, this Journal, vol. 3(1937), pp. 544–548.

[3] C. C. MacDuffee, Monatshefte für Mathematik und Physik, vol. 48(1939), pp. 293–313. See p. 294, (2, 6) transposed. The result is due to Frobenius.

a subset of the matrices $R_i\bar{R}_j$ which are linearly independent and in terms of which every product $R_i\bar{R}_j$ is representable. Then $D_1, D_2, \cdots, D_p$ form an $\mathfrak{F}$-basis for $\mathfrak{D}$.

THEOREM 1. *The algebra $\mathfrak{D}$ is a representation of the first derived algebra of $\mathfrak{A}$.*

Since $\epsilon_1 = 1$, both R_1 and $\bar{R}_1$ are equal to the identity matrix of order n. Then $\mathfrak{D}$ contains $R_iI = R_i$, and also $I\bar{R}_j = \bar{R}_j$, so $\mathfrak{D} \supseteq \mathfrak{A}$ and $\mathfrak{D} \supseteq \bar{\mathfrak{A}}$. If $\mathfrak{D}'$ is any algebra which includes every R_i and every $\bar{R}_j$, then $\mathfrak{D}'$ must contain all products $R_i\bar{R}_j$, so that $\mathfrak{D}' \supseteq \mathfrak{D}$. That is, $\mathfrak{D}$ is the first derived algebra of $\mathfrak{A}$.

Evidently, from (2), every number of $\mathfrak{A}$ is commutative in $\mathfrak{D}$ with every number of $\bar{\mathfrak{A}}$.

Since $\mathfrak{A} \subseteq \mathfrak{D} \subseteq \mathfrak{A} \times \bar{\mathfrak{A}}$, it follows that

$$n \leqq p \leqq n^2,$$

the lower value being attained when $\mathfrak{A}$ is commutative. That the upper value for p is sometimes reached follows from a result of Jacobson.[4] An algebra $\mathfrak{A}$ over $\mathfrak{F}$ is called normal simple if the algebra obtained from $\mathfrak{A}$ by extending $\mathfrak{F}$ to its algebraic closure is simple, and Jacobson proved that if $\mathfrak{A}$ is normal simple, then $\mathfrak{D} = \mathfrak{A} \times \bar{\mathfrak{A}}$, and that $\mathfrak{D}$ is a complete matric algebra. Since a complete matric algebra is normal simple, $p = n^2$ when $\mathfrak{A}$ is complete matric.

Jacobson also proved that if $\mathfrak{A}$ is semisimple, i. e.,

$$\mathfrak{A} = \mathfrak{A}_1 \dotplus \mathfrak{A}_2 \dotplus \cdots \dotplus \mathfrak{A}_k,$$

where each $\mathfrak{A}_i$ is simple, then

$$\mathfrak{D} = \mathfrak{D}_1 \dotplus \mathfrak{D}_2 \dotplus \cdots \dotplus \mathfrak{D}_k,$$

where $\mathfrak{D}_i$ is simple and is the first derived algebra of $\mathfrak{A}_i$.

We shall prove

THEOREM 2. *If $\mathfrak{A} = \mathfrak{A}_1 \times \mathfrak{A}_2 \times \cdots \times \mathfrak{A}_k$, then $\mathfrak{D} = \mathfrak{D}_1 \times \mathfrak{D}_2 \times \cdots \times \mathfrak{D}_k$, where $\mathfrak{D}_i$ is the first derived algebra of $\mathfrak{A}_i$.*

Take $k = 2$, and let $\mathfrak{A}_1$ have the basis

$$\epsilon_1, \epsilon_2, \cdots, \epsilon_g, \qquad \epsilon_i\epsilon_j = \sum_k c_{ijk}\epsilon_k,$$

and $\mathfrak{A}_2$ the basis

$$\eta_1, \eta_2, \cdots, \eta_h, \qquad \eta_i\eta_j = \sum_k c'_{ijk}\eta_k.$$

Then $\mathfrak{A}_1 \times \mathfrak{A}_2$ has the basis

$$\epsilon_1\eta_1, \epsilon_2\eta_1, \cdots, \epsilon_g\eta_1, \epsilon_1\eta_2, \cdots, \epsilon_g\eta_2, \cdots, \epsilon_g\eta_h,$$

where $\epsilon_i\eta_j = \eta_j\epsilon_i$. It is clear that if

$$\epsilon_i\eta_k\epsilon_j\eta_l = \sum_{p,q} c_{ij,kl,pq}\epsilon_p\eta_q,$$

[4] N. Jacobson, loc. cit., p. 547.

the constants of multiplication of $\mathfrak{A}_1 \times \mathfrak{A}_2$ are

$$c_{ij,kl,pq} = c_{ikp}c'_{jlq}\,.$$

Hence

$$R(\epsilon_i\eta_j) = (c_{ij,s_1s_2,r_1r_2}) = (c_{is_1r_1}c'_{js_2r_2}) = R(\epsilon_i) \times R(\eta_j).$$

Here $R(\epsilon_i\eta_j)$ is the first regular representation of $\epsilon_i\eta_j$ in $\mathfrak{A}$, while $R(\epsilon_i)$ [or $R(\eta_j)$] is the first regular representation of ϵ_1 in $\mathfrak{A}_1$ [or of η_j in $\mathfrak{A}_2$], and $\times$ denotes the direct product.[5] Hence

$$R(\epsilon_i\eta_j)\bar{R}(\epsilon_k\eta_l) = (R(\epsilon_i) \times R(\eta_j))(\bar{R}(\epsilon_k) \times \bar{R}(\eta_l)) = R(\epsilon_i)\bar{R}(\epsilon_k) \times R(\eta_j)\bar{R}(\eta_l).$$

Clearly every number of $\mathfrak{D}$ is in $\mathfrak{D}_1 \times \mathfrak{D}_2$, and conversely.

The extension to any finite k is immediate.

3. **Analytic functions.** As before, let $\mathfrak{A}$ be a finite linear associative algebra of order n over a field $\mathfrak{F}$, with unit element. We shall now assume that $\mathfrak{F}$ is a field with n independent derivations. We recall that a derivation D is an endomorphism, or mapping of $\mathfrak{F}$ onto itself, such that

$$D(x \pm y) = Dx \pm Dy, \qquad D(xy) = xDy + yDx$$

for every x and y in $\mathfrak{F}$. We shall assume the existence of n derivations

$$D_1, D_2, \cdots, D_n$$

and n numbers $x_1, x_2, \cdots, x_n$ of $\mathfrak{F}$ such that

$$D_ix_j = \delta_{ij},$$

where δ_{ij} is Kronecker's delta.

In particular $\mathfrak{F}$ may be the transcendental extension $\mathfrak{F} = \varphi(x_1, x_2, \cdots, x_n)$ of a field φ, so that $\mathfrak{F}$ consists of all rational functions of the n indeterminates $x_1, x_2, \cdots, x_n$ with coefficients in φ. Or, more generally, $\mathfrak{F}$ may be a function field in the n independent variables $x_1, x_2, \cdots, x_n$, where the functions are restricted in various ways. We shall adopt the usual notation, namely,

$$D_i = \frac{\partial}{\partial x_i},$$

and assume that the first partial derivatives exist in $\mathfrak{F}$. No matter in what way the word function may be defined for $\mathfrak{F}$, it must be true that, if $y_1, y_2, \cdots, y_n$ are functions of $x_1, x_2, \cdots, x_n$, and $z_1, z_2, \cdots, z_n$ are functions of $y_1, y_2, \cdots, y_n$, then

$$\frac{\partial z_i}{\partial x_j} = \sum_{h=1}^{n} \frac{\partial z_i}{\partial x_h}\frac{\partial x_h}{\partial x_j} \qquad (i, j = 1, 2, \cdots, n). \tag{3}$$

[5] C. C. MacDuffee, *The Theory of Matrices*, Berlin, 1933, p. 82.

For some theorems it will be necessary to assume also that the second partial derivatives exist in $\mathfrak{F}$.

The matrix

$$\left[\frac{\partial y_r}{\partial x_s}\right] \qquad (r, s = 1, 2, \cdots, n),$$

where r is the row-index and s the column-index, is called the Jacobian matrix of $y_1, y_2, \cdots, y_n$ with respect to $x_1, x_2, \cdots, x_n$.

If $y_1, y_2, \cdots, y_n$ are functions of $x_1, x_2, \cdots, x_n$, we shall say that the number

$$\eta = y_1\epsilon_1 + y_2\epsilon_2 + \cdots + y_n\epsilon_n$$

is a function of the number

$$\xi = x_1\epsilon_1 + x_2\epsilon_2 + \cdots + x_n\epsilon_n,$$

and we shall call the Jacobian matrix

$$\left[\frac{\partial y_r}{\partial x_s}\right] = \frac{d\eta}{d\xi}$$

the derivative of η with respect to ξ. Thus the derivative always exists in the total matric algebra $\mathfrak{M}$ of order n^2 in which $\mathfrak{A}$ is imbedded.

If $d\eta/d\xi$ is in the first derived algebra $\mathfrak{D}$ of $\mathfrak{A}$, we shall call η an analytic function of ξ.

Certain numbers of $\mathfrak{A}$ will have derivatives which lie in $\mathfrak{A}$, but it must be emphasized that these numbers do not in general form a closed set under multiplication. Clearly $d\eta/d\xi$ will lie in $\mathfrak{A}$ if and only if $d\eta/d\xi$ is an R-matrix. It has been shown[6] that a matrix is an R-matrix if and only if it is commutative with every $\bar{R}(\alpha)$. We then have

Theorem 3. *The derivative $d\eta/d\xi$ lies in $\mathfrak{A}$ if and only if $d\eta/d\xi$ is commutative with $\bar{R}(\alpha)$ for every α in $\mathfrak{A}$.*

Let us see what this definition of analyticity means when $\mathfrak{A}$ is the complex function field, and $\mathfrak{F}$ the field of real functions of x_1, x_2 whose partials exist. Then $\epsilon_1 = 1$, $\epsilon_2 = i$,

$$R_1 = \bar{R}_1 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \qquad R_2 = \bar{R}_2 = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}.$$

Since $\mathfrak{A}$ is commutative, $\mathfrak{A} = \mathfrak{D}$. If

$$\xi = x_1 + x_2 i, \qquad \eta = y_1 + y_2 i,$$

[6] MacDuffee, loc. cit. (footnote 2), pp. 294–295.

then according to our definition η is an analytic function of ξ if real functions A and B exist such that

$$\frac{d\eta}{d\xi} = \begin{bmatrix} \dfrac{\partial y_1}{\partial x_1} & \dfrac{\partial y_1}{\partial x_2} \\ \dfrac{\partial y_2}{\partial x_1} & \dfrac{\partial y_2}{\partial x_2} \end{bmatrix} = AR_1 + BR_2 = \begin{bmatrix} A & -B \\ B & A \end{bmatrix}.$$

This means that

$$A = \frac{\partial y_1}{\partial x_1} = \frac{\partial y_2}{\partial x_2}, \qquad B = \frac{\partial y_2}{\partial x_1} = -\frac{\partial y_1}{\partial x_2}.$$

These are the familiar Cauchy-Riemann differential equations. If they are fulfilled,

$$\frac{d\eta}{d\xi} = \frac{\partial y_1}{\partial x_1}R_1 + \frac{\partial y_2}{\partial x_1}R_2 \leftrightarrow \frac{\partial y_1}{\partial x_1} + \frac{\partial y_2}{\partial x_1}i.$$

The algebra with the multiplication table

(4)

	ϵ_1	ϵ_2	ϵ_3
ϵ_1	ϵ_1	ϵ_2	ϵ_3
ϵ_2	ϵ_2	ϵ_1	ϵ_3
ϵ_3	ϵ_3	$-\epsilon_3$	0

is associative but not commutative, and has the unit element ϵ_1. Let $\mathfrak{F}$ be the field of all real differentiable functions of x_1, x_2, x_3. If $\xi = x_1\epsilon_1 + x_2\epsilon_2 + x_3\epsilon_3$, then

$$R(\xi) = \begin{bmatrix} x_1 & x_2 & 0 \\ x_2 & x_1 & 0 \\ x_3 & -x_3 & x_1 + x_2 \end{bmatrix}, \qquad \bar{R}(\xi) = \begin{bmatrix} x_1 & x_2 & 0 \\ x_2 & x_1 & 0 \\ x_3 & x_3 & x_1 - x_2 \end{bmatrix}.$$

As a basis for the derived algebra $\mathfrak{D}$, we may take

$$D_1 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \qquad D_2 = \begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \qquad D_3 = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 1 & 0 & 0 \end{bmatrix},$$

$$D_4 = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}, \qquad D_5 = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$

Thus $\eta = y_1\epsilon_1 + y_2\epsilon_2 + y_3\epsilon_3$ will be an analytic function of ξ if and only if functions $A_1, A_2, \cdots, A_5$ of $\mathfrak{F}$ exist such that

$$\frac{d\eta}{d\xi} = \begin{bmatrix} \dfrac{\partial y_1}{\partial x_1} & \dfrac{\partial y_1}{\partial x_2} & \dfrac{\partial y_1}{\partial x_3} \\ \dfrac{\partial y_2}{\partial x_1} & \dfrac{\partial y_2}{\partial x_2} & \dfrac{\partial y_2}{\partial x_3} \\ \dfrac{\partial y_3}{\partial x_1} & \dfrac{\partial y_3}{\partial x_2} & \dfrac{\partial y_3}{\partial x_3} \end{bmatrix} = \begin{bmatrix} A_1 & A_2 & 0 \\ A_2 & A_1 & 0 \\ A_3 & A_4 & A_5 \end{bmatrix}.$$

This means that

$$\frac{\partial y_1}{\partial x_3} = \frac{\partial y_2}{\partial x_3} = 0, \qquad \frac{\partial y_1}{\partial x_1} = \frac{\partial y_2}{\partial x_2}, \qquad \frac{\partial y_1}{\partial x_2} = \frac{\partial y_2}{\partial x_1}.$$

These are the generalized Cauchy-Riemann equations for this algebra.

By looking at the form of $R(\xi)$, it is evident what further conditions are necessary in order that $d\eta/d\xi$ may lie in $\mathfrak{A}$. These conditions are

$$\frac{\partial y_3}{\partial x_1} = -\frac{\partial y_3}{\partial x_2}, \qquad \frac{\partial y_3}{\partial x_3} = \frac{\partial y_1}{\partial x_1} + \frac{\partial y_1}{\partial x_2}.$$

4. **Properties of analytic functions.** If according to our definition

$$\zeta = z_1\epsilon_1 + z_2\epsilon_2 + \cdots + z_n\epsilon_n$$

is a function of η, and if η is a function of ξ, then ζ is a function of ξ. Also, from (3),

$$\left[\frac{\partial z_r}{\partial x_s}\right] = \left[\sum_{h=1}^{n} \frac{\partial z_r}{\partial y_h}\frac{\partial y_h}{\partial x_s}\right] = \left[\frac{\partial z_r}{\partial y_s}\right]\left[\frac{\partial y_r}{\partial x_s}\right],$$

so that

$$\frac{d\zeta}{d\xi} = \frac{d\zeta}{d\eta}\frac{d\eta}{d\xi}.$$

If η is an analytic function of ξ, then $d\eta/d\xi$ is in $\mathfrak{D}$; if also ζ is an analytic function of η, then $d\zeta/d\eta$ is in $\mathfrak{D}$, and therefore $d\zeta/d\xi$ is in $\mathfrak{D}$. We have

THEOREM 4. *If ζ is an analytic function of η, and if η is an analytic function of ξ, then ζ is an analytic function of ξ, and*

$$\frac{d\zeta}{d\xi} = \frac{d\zeta}{d\eta}\frac{d\eta}{d\xi}.$$

When $\zeta = \xi$, we have the familiar formula for the derivative of the inverse function, namely,

$$\frac{d\xi}{d\eta}\cdot\frac{d\eta}{d\xi} = 1.$$

If $\zeta = \eta_1 + \eta_2$, where η_1 and η_2 are functions of ξ, it is clear that

$$\left[\frac{\partial z_r}{\partial x_s}\right] = \left[\frac{\partial y_{1r}}{\partial x_s}\right] + \left[\frac{\partial y_{2r}}{\partial x_s}\right],$$

and thus we have the following theorem.

THEOREM 5. *The derivative of the sum of two functions is equal to the sum of their derivatives. The sum of two analytic functions is analytic.*

Let us now consider the product of two functions

$$\eta = y_1\epsilon_1 + \cdots + y_n\epsilon_n, \qquad \zeta = z_1\epsilon_1 + \cdots + z_n\epsilon_n.$$

Then

$$\eta\zeta = \sum_{i,j} y_i z_j \epsilon_i \epsilon_j = \sum_{i,j,k} y_i z_j c_{ijk} \epsilon_k .$$

According to our definition,

$$\frac{d(\eta\zeta)}{d\xi} = \left[\frac{\partial}{\partial x_s} \sum_{i,j} y_i z_j c_{ijr}\right] = \left[\sum_{i,j} y_i c_{ijr} \frac{\partial z_j}{\partial x_s}\right] + \left[\sum_{i,j} z_j c_{ijr} \frac{\partial y_i}{\partial x_s}\right]$$

$$= R(\eta)\frac{d\zeta}{d\xi} + \bar{R}(\zeta)\frac{d\eta}{d\xi}.$$

Now both $R(\eta)$ and $\bar{R}(\zeta)$ are in $\mathfrak{D}$; hence if both $d\zeta/d\xi$ and $d\eta/d\xi$ are in $\mathfrak{D}$, so is $d(\eta\zeta)/d\xi$. We have proved

THEOREM 6. *For any two functions η and ζ,*

$$\frac{d(\eta\zeta)}{d\xi} = \eta\frac{d\zeta}{d\xi} + \bar{\zeta}\frac{d\eta}{d\xi},$$

where $\bar{\zeta}$ is the correspondent of ζ under the anti-isomorphism $\mathfrak{A} \cong \bar{\mathfrak{A}}$. The product of two analytic functions is analytic.

Theorems 4, 5 and 6 furnish a basis for finding the derivatives of all functions which can be built up by the rational operations. Thus

$$\frac{d}{d\xi}\eta^r = [\eta^{r-1} + \bar{\eta}\eta^{r-2} + \bar{\eta}^2\eta^{r-3} + \cdots + \bar{\eta}^{r-1}]\frac{d\eta}{d\xi}.$$

If η and ζ are reciprocals so that $\eta\zeta = 1$, then

$$\eta\frac{d\zeta}{d\xi} + \bar{\zeta}\frac{d\eta}{d\xi} = 0,$$

$$\frac{d\eta}{d\xi} = -\bar{\zeta}^{-1}\eta\frac{d\zeta}{d\xi} = -(\zeta\bar{\zeta})^{-1}\frac{d\zeta}{d\xi}.$$

All these results become well-known formulas when $\mathfrak{A}$ is commutative.

5. **Integration.** In this section we must assume that the second partial derivatives exist in $\mathfrak{F}$.

As Scheffers pointed out, not every number of $\mathfrak{A}$ can have an antiderivative unless $\mathfrak{A}$ is commutative. However, certain functions in $\mathfrak{A}$ can have antiderivatives in $\mathfrak{A}$, and we shall find necessary and sufficient conditions that this may be so.

Let us suppose that

$$\zeta = z_1 R_1 + \cdots + z_n R_n = \frac{d\eta}{d\xi}.$$

That is,

$$\frac{\partial y_r}{\partial x_s} = \sum_i z_i c_{isr} \qquad (r, s = 1, 2, \cdots, n).$$

When ζ is given, the functions $y_1, y_2, \cdots, y_n$ exist if and only if the integrability conditions

$$\frac{\partial^2 y_r}{\partial x_t \partial x_s} = \frac{\partial^2 y_r}{\partial x_s \partial x_t} \qquad (r, s, t = 1, 2, \cdots, n)$$

are satisfied. This is the same thing as saying that

$$\frac{\partial^2 y_h}{\partial x_r \partial x_s} = \sum_i \frac{\partial z_i}{\partial x_s} c_{irh} \qquad (r, s, h = 1, 2, \cdots, n)$$

shall be invariant under the interchange of r and s. In matric notation, this means that the matrix[7]

$$Q_h^T \frac{d\zeta}{d\xi}, \qquad Q_h^T = (c_{srh}), \tag{5}$$

shall be symmetric for every h. We have

THEOREM 7. *A necessary and sufficient condition in order that ζ may have an antiderivative in $\mathfrak{A}$ is that the matrix*

$$Q^T(\alpha) \frac{d\zeta}{d\xi}$$

shall be symmetric for every number α in $\mathfrak{A}$.

To see that there exist integrable functions in non-commutative algebras, consider algebra (4), and take

$$\zeta = (x_1 - x_2)\epsilon_1 + (-x_1 + x_2)\epsilon_2 + (-x_1 + x_2)\epsilon_3,$$
$$\alpha = a_1\epsilon_1 + a_2\epsilon_2 + a_3\epsilon_3.$$

We have

$$Q^T(\alpha)\frac{d\zeta}{d\xi} = \begin{bmatrix} a_1 & a_2 & a_3 \\ a_2 & a_1 & -a_3 \\ a_3 & a_3 & 0 \end{bmatrix} \cdot \begin{bmatrix} 1 & -1 & 0 \\ -1 & 1 & 0 \\ -1 & 1 & 0 \end{bmatrix} = \begin{bmatrix} a_1 - a_2 - a_3 & -a_1 + a_2 + a_3 & 0 \\ -a_1 + a_2 + a_3 & a_1 - a_2 - a_3 & 0 \\ 0 & 0 & 0 \end{bmatrix},$$

which is symmetric, so ζ is integrable.

The integration can be carried out in conventional notation if we define

$$d\xi = dx_1\epsilon_1 + dx_2\epsilon_2 + \cdots + dx_n\epsilon_n. \tag{6}$$

Then

$$\int \zeta d\xi = \int \sum_{i,j,k} z_i \, dx_j \, c_{ijk} \, \epsilon_k = \sum_k \left[\sum_j \int \sum_i z_i c_{ijk} \, dx_j \right] \epsilon_k.$$

The condition for exactness, namely,

$$\sum_i \frac{\partial z_i}{\partial x_h} c_{ijk} = \sum_i \frac{\partial z_i}{\partial x_j} c_{ihk} \qquad (h, j, k = 1, 2, \cdots, n),$$

is again our condition (5).

[7] The matrix $Q(\alpha) = \Sigma a_i Q_i$ was called by Frobenius the *parastrophic matrix* of α. MacDuffee, loc. cit. (footnote 2), p. 296.

It seems easier, however, to employ the matric notation. In our example,

$$R(\zeta) = \begin{bmatrix} x_1 - x_2 & -x_1 + x_2 & 0 \\ -x_1 + x_2 & x_1 - x_2 & 0 \\ -x_1 + x_2 & x_1 - x_2 & 0 \end{bmatrix} = \left(\frac{\partial y_r}{\partial x_s}\right),$$

so that

$$\eta = \int \zeta \, d\xi = y_1 \epsilon_1 + y_2 \epsilon_2 + y_3 \epsilon_3,$$

where

$$\begin{aligned} y_1 &= \tfrac{1}{2}x_1^2 - x_1 x_2 + \tfrac{1}{2}x_2^2 + f_1(x_3), \\ y_2 &= -\tfrac{1}{2}x_1^2 + x_1 x_2 - \tfrac{1}{2}x_2^2 + f_2(x_3), \\ y_3 &= -\tfrac{1}{2}x_1^2 + x_1 x_2 - \tfrac{1}{2}x_2^2 + f_3(x_3), \end{aligned}$$

the f's being arbitrary functions.

Let us suppose that ζ is integrable, so that for every t

$$Q_t^T \frac{d\zeta}{d\xi}$$

is symmetric. That is,

$$\sum_i c_{irt} \frac{\partial z_i}{\partial x_s} = \sum_i c_{ist} \frac{\partial z_i}{\partial x_r} \qquad (r, s, t = 1, 2, \cdots, n).$$

After the substitution $r \to s$, $s \to t$, $t \to r$, this becomes

$$\sum_i c_{isr} \frac{\partial z_i}{\partial x_t} = \sum_i c_{itr} \frac{\partial z_i}{\partial x_s} \qquad (r, s, t = 1, 2, \cdots, n).$$

In matric notation this is

$$\sum_i \frac{\partial z_i}{\partial x_t} R_i = \bar{R}_t \frac{d\zeta}{d\xi}. \tag{7}$$

By $\partial\zeta/\partial x_t$ we shall understand the R-matrix of

$$\frac{\partial z_1}{\partial x_t} \epsilon_1 + \frac{\partial z_2}{\partial x_t} \epsilon_2 + \cdots + \frac{\partial z_n}{\partial x_t} \epsilon_n,$$

or what is the same thing, the matrix obtained from $R(\zeta)$ by operating upon every element with $\partial/\partial x_t$. Then (7) may be written

$$\frac{\partial \zeta}{\partial x_t} = \bar{R}_t \frac{d\zeta}{d\xi} \qquad (i = 1, 2, \cdots, n). \tag{8}$$

Since every step is reversible, it is clear that (8) implies (5). We shall call

$$P_a = \sum_t a_t \frac{\partial}{\partial x_t}$$

a *polar operator*. We have now proved

THEOREM 8. *A necessary and sufficient condition in order that ζ have an antiderivative in $\mathfrak{A}$ is that, for every number α in $\mathfrak{A}$,*

$$P_\alpha \zeta = \bar{R}(\alpha) \frac{d\zeta}{d\xi}.$$

In particular, if ζ is integrable,

$$\frac{d\zeta}{d\xi} = \frac{\partial \zeta}{\partial x_1}.$$

This last statement follows from the fact that if ϵ_1 is a unit element, $\bar{R}(\epsilon_1)$ is the identity matrix, and $P_{\epsilon_1} = \partial/\partial x_1$. The validity of the last equation in the theorem is not alone sufficient that ζ be integrable.

6. **Commutative algebras.** When $\mathfrak{A}$ is commutative, $\mathfrak{D} = \mathfrak{A}$, and the entire theory becomes much simpler. In particular, the fact that every analytic function in $\mathfrak{A}$ has a derivative in $\mathfrak{A}$ is a trivial consequence of the definition of analytic function. The derivative $d\eta/d\xi$ is now an R-matrix.

The following theorem is less trivial.

THEOREM 9. *A necessary and sufficient condition that ζ be integrable in a commutative algebra $\mathfrak{A}$ is that ζ be analytic.*

We shall show that when $\mathfrak{A}$ is commutative, the criterion of Theorem 7 is fulfilled for every analytic function ζ. Since the derivative of ζ exists in $\mathfrak{A}$, we can write

$$\frac{d\zeta}{d\xi} = w_1 R_1 + w_2 R_2 + \cdots + w_n R_n.$$

If $\alpha = a_1\epsilon_1 + a_2\epsilon_2 + \cdots + a_n\epsilon_n$, then

$$Q^T(\alpha) \frac{d\zeta}{d\xi} = \sum_{i,j} a_i w_j Q_i^T R_j.$$

Now it is true that[8]

$$Q_i^T R_j = \sum_k c_{jki} Q_k^T \qquad (i, j = 1, 2, \cdots, n).$$

Hence $Q^T(\alpha)\, d\zeta/d\xi$ is a linear combination of matrices $Q_i^T = (c_{sri})$, each of which is symmetric when $\mathfrak{A}$ is commutative. Hence if ζ is analytic, it is integrable.

Conversely, if ζ is integrable, it is true from Theorem 8 that

$$\frac{d\zeta}{d\xi} = \frac{\partial \zeta}{\partial x_1} = \frac{\partial z_1}{\partial x_1} R_1 + \frac{\partial z_2}{\partial x_1} R_2 + \cdots + \frac{\partial z_n}{\partial x_1} R_n,$$

which is surely in $\mathfrak{A}$. Hence ζ is analytic.

[8] MacDuffee, loc. cit. (footnote 2), p. 294, (2, 8) transposed.

7. Relation to the literature. The paper of Scheffers [50] which was mentioned in §1 dealt mainly with commutative algebras, and is properly included in that of Hausdorff [26]. Of all definitions of analytic function to be found in the literature, that of Hausdorff is the most general. He called η an analytic function of ξ if functions φ_{ij} exist such that

$$d\eta = \sum_{i,j} \frac{\partial y_i}{\partial x_j} dx_j \epsilon_i = \sum_{i,j} \varphi_{ij} \epsilon_i \, d\xi \, \epsilon_j .$$

Using (6) and equating coefficients of $dx_j \epsilon_i$, we have

$$\frac{\partial y_r}{\partial x_s} = \sum_{i,j,k} \varphi_{ij} c_{isk} c_{kjr} \qquad (r, s = 1, 2, \cdots, n).$$

In matric notation this is

$$\left[\frac{\partial y_r}{\partial x_s}\right] = \sum_{i,j} \varphi_{ij} \bar{R}_j R_i = \sum_{i,j} \varphi_{ij} R_i \bar{R}_j .$$

This is precisely our definition of §3. From his definition, Hausdorff was able to obtain generalized Cauchy-Riemann differential equations, but he did not succeed in obtaining a definition of derivative to go with his definition of analytic function.

Improvements and simplifications in the Scheffers-Hausdorff theory were made by Ketchum [31] and Ringleb [49].

Spampinato [68] called η "totally derivable on the left" if there exists a ζ in $\mathfrak{A}$ such that

$$d\eta = \zeta \, d\xi, \tag{9}$$

where $d\eta$ and $d\xi$ are defined as in (6). But

$$\begin{aligned} d\eta &= \sum_i dy_i \epsilon_i = \sum_{i,j} \frac{\partial y_i}{\partial x_j} dx_j \epsilon_i \\ &= \sum_{i,j} z_i \epsilon_i \, dx_j \epsilon_j = \sum_{i,j,k} z_i \, dx_j c_{ijk} \epsilon_k . \end{aligned}$$

Hence (9) is equivalent to the equations

$$\frac{\partial y_r}{\partial x_s} = \sum_i z_i c_{isr} \qquad (r, s = 1, 2, \cdots, n).$$

In our notation, this is

$$\frac{d\eta}{d\xi} = R(\zeta).$$

Thus $d\eta/d\xi$ is in $\mathfrak{A}$. **Similarly η is "totally derivable on the right" if there exists a ζ' such that**

$$d\eta = d\xi \, \zeta'.$$

In our notation this means

$$\frac{d\eta}{d\xi} = \bar{R}(\zeta').$$

This is in $\mathfrak{A}$. These definitions are more restrictive than ours. The principal contribution of Spampinato in connection with our paper is his pointing out that in the commutative case the Jacobian $(\partial y_r/\partial x_s)$ is the regular representation of the derivative.

Bibliography of Function Theory in Linear Algebras

1. V. Amato, *Sulle definizioni di funzioni di matrici nel corpo complesso*, Bollettino della Unione Matematica Italiana, Bologna, vol. 12(1933), pp. 243–249.
2. G. Andreoli, *Equazioni e sistemi differenziali, operatori funzionali, funzioni nelle algebra*, I, II, Acta Societatis Gioeniae Catinensis Naturalium Scientiarum, Catania, vol. 19(1933), mem. 12; vol. 20(1934), mem. 1.
3. G. Andreoli, *Sulle funzioni di composizione di matrici (funzioni isogene)*, Atti della Reale Accademia delle Scienze fisiche e matematiche, Napoli, vol. 20(1935), mem. 10, pp. 1–30.
4. L. Autonne, *Sur les propriétés qui, pour les fonctions d'une variable hypercomplexe, correspondent à la monogénéité*, Journal de Mathématiques pures et appliquées, (6), vol. 3(1907), pp. 53–104.
5. H. A. von Beckh-Widmanstetter, *Läszt sich die Eigenschaft der analytischen Funktionen einer gemeinen komplexen Veränderlichen, Potentiale als Bestandteile zu liefern, auf Zahlsysteme mit drei Einheiten verallgemeinern?*, Monatshefte für Mathematik und Physik, vol. 23(1912), pp. 257–260.
6. G. Birkhoff, *On product integration*, Journal of Mathematics and Physics, Massachusetts Institute of Technology, vol. 16(1937), pp. 104–132.
7. C. Carbonaro, *Sulle funzioni di una variabile biduale totalmente derivabili*, Atti della Reale Accademia Nazionale dei Lincei, Rendiconti, (6), vol. 23(1936), pp. 839–845.
8. Giosuè Catalano, *Sul luogo dei punti regolari di una funzione analitica tali che i corrispondenti cerchi di convergenza contengano un punto assegnato*, Atti dell'Accademia Gioenia di Scienze naturali in Catania, (6), vol. 1(1936), mem. 7, pp. 1–5.
9. S. Cherubino, *Sulle radici caratteristiche delle funzioni olomorfe di matrici*, Atti della Reale Accademia Nazionale dei Lincei, Rendiconti, (6), vol. 23(1936), pp. 846–849.
10. S. Cherubino, *Funzioni olomorfe di matrici*, Annali della Reale Scuola Normale Superiore di Pisa, (2), vol. 6(1937), pp. 41–70.
11. S. Cherubino, *Applicazione delle funzioni olomorfe di matrici ai sistemi di equazioni differenziali lineari*, I, II, Atti della Reale Accademia Nazionale dei Lincei, Rendiconti, (6), vol. 25(1937), pp. 541–547, 686–690.
12. S. Cherubino, *Integrale di Volterra e funzioni olomorfe di matrici*, Annali della Reale Scuola Normale Superiore di Pisa, (2), vol. 7(1938), pp. 313–328.
13. Michele Cipolla, *Sulle matrici espressioni analitiche di un'altra*, Rendiconti del Circolo Matematico di Palermo, vol. 56(1932), pp. 144–154.
14. U. Cisotti, *Deduzioni differenziali dalla definizione di vettori reciproci*, Atti della Reale Accademia Nazionale dei Lincei, Rendiconti, (6), vol. 18(1933), pp. 255–260, 331–334, 420–424, 469–472; vol. 19(1934), pp. 3–8.
15. G. Scorza Dragoni, *Sulle funzioni olomorfe di una variabile bicomplessa*, Reale Accademia d'Italia, Memorie della classe di scienze fisiche, matematiche e naturali, vol. 5(1934), pp. 597–665.

16. Nelson Dunford, *Integration in general analysis*, Transactions of the American Mathematical Society, vol. 37(1935), pp. 441–453.
17. M. Fréchet, *La notion de différentielle dans l'analyse générale*, Annales Scientifiques de l'Ecole Normale Supérieure, (3), vol. 42(1925), pp. 293–323.
18. R. Fueter, *Die Theorie der regulären Funktionen einer Quaternionenvariablen*, Oslo Comptes Rendus du Congrès International des Mathématiciens, Oslo, 1936, pp. 75–91.
19. R. Fueter, *Zur Theorie der regulären Funktionen einer Quaternionenvariablen*, Monatshefte für Mathematik und Physik, vol. 43(1936), pp. 69–74.
20. R. Fueter, *Über die analytische Darstellung der regulären Funktionen einer Quaternionenvariablen*, Commentarii Mathematici Helvetici, vol. 8(1936), pp. 371–378.
21. M. Futagawa, *On the theory of functions of a quaternary variable*, I, II, Tôhoku Mathematical Journal, vol. 29(1928), pp. 175–222; vol. 35(1932), pp. 69–120.
22. G. Giorgi, *Nuove osservazioni sulle funzioni delle matrici*, Atti della Reale Accademia Nazionale dei Lincei, Rendiconti, (6), vol. 8(1928), pp. 3–8.
23. E. Göllnitz, *Beiträge zur Quaternionentheorie*, Akademie der Wissenschaften in Wien, mathematisch-naturwissenschaftliche Klasse, Sitzungsberichte, vol. 137(1928), pp. 157–188.
24. E. Göllnitz, *Über die Quaternionenfunktionen* log r *und* arc tang r, Akademie der Wissenschaften in Wien, mathematisch-naturwissenschaftliche Klasse, Sitzungsberichte, vol. 137(1928), pp. 351–362.
25. L. M. Graves, *Topics in the functional calculus*, Bulletin of the American Mathematical Society, vol. 41(1935), pp. 641–662.
26. F. Hausdorff, *Zur Theorie der Systeme complexer Zahlen*, Berichte über die Verhandlungen der Sächsischen Akademie der Wissenschaften zu Leipzig, mathematisch-physikalische Klasse, vol. 52(1900), pp. 43–61.
27. K. Hensel, *Über Potenzreihen von Matrizen*, Journal für die reine und angewandte Mathematik, vol. 155(1926), pp. 107–110.
28. B. Hostinský, *Sur la notion d'intégrale d'une substitution linéaire*, Mathematica (Cluj), vol. 10(1935), pp. 191–193.
29. B. Hostinský, *Sur l'intégration des substitutions linéaires*, Atti della Reale Accademia Nazionale dei Lincei, Rendiconti, (6), vol. 22(1935), pp. 221–225.
30. A. Kawaguchi, *Displacements in a manifold of matrices*, I, II, Proceedings of the Imperial Academy of Japan, vol. 11(1935), pp. 39–42, 168–170.
31. P. W. Ketchum, *Analytic functions of hypercomplex variables*, Transactions of the American Mathematical Society, vol. 30(1928), pp. 641–667.
32. P. W. Ketchum, *Solution of partial differential equations by means of hypervariables*, American Journal of Mathematics, vol. 54(1932), pp. 253–264.
33. P. W. Ketchum and Ted Martin, *Polygenic functions of hypercomplex variables*, Bulletin of the American Mathematical Society, vol. 38(1932), pp. 66–72.
34. J. A. Lappo-Danilevskij, *Mémoires sur la théorie des systèmes des équations différentielles linéaires*, Akademia nauk Petrograd Fizoko-matematicheskii Institut Bulletin, vol. 6(1934), pp. 1–256.
35. A. Loewy, *Anschauliche Interpretation eines linearen homogenen Differentialsystems*, Compositio Mathematica, vol. 1(1934), pp. 188–192.
36. A. Loewy, *Über ein Integralsystem mit Stieltjesschen Integralen, seine anschauliche Interpretation und den Infinitesimalkalkül für Matrizen*, Compositio Mathematica, vol. 2(1935), pp. 417–423.
37. S. Martis, *Calcolo del logaritmo di una matrice di secondo ordine*, Atti della Reale Accademia Nazionale dei Lincei, Rendiconti, (6), vol. 8(1921), pp. 474–480.
38. A. D. Michal and A. H. Clifford, *Fonctions analytiques implicites dans des espaces vectoriels abstraits*, Comptes Rendus hebdomadaires des Séances de l'Académie des Sciences, Paris, vol. 197(1933), pp. 735–737.

39. A. D. Michal and R. S. Martin, *Some expansions in vector space*, Journal de Mathématiques pures et appliquées, (9), vol. 13(1934), pp. 69–91.
40. A. K. Mitchell, *On a matrix differential operator*, Bulletin of the American Mathematical Society, vol. 38(1932), pp. 251–254.
41. Gr. C. Moisil, *Sur le calcul différentiel absolu des variétés plongées dans l'espace fonctionnel*, Annales Scientifiques de l'Université de Jassy, vol. 16(1931), pp. 375–382.
42. Gr. C. Moisil, *Sur l'intégration des matrices*, Comptes Rendus hebdomadaires des Séances de l'Académie des Sciences, Paris, vol. 198(1932), pp. 456–458.
43. Gr. C. Moisil and N. Théodoresco, *Fonctions holomorphes dans l'espace*, Mathematica (Cluj), vol. 5(1931), pp. 142–159.
44. K. Morinaga, *An extension of parallel displacement by matrices*, Journal of Science of the Hiroshima University Series A (Mathematics, Physics, Chemistry), vol. 4(1934), pp. 127–140.
45. M. Nagumo, *Einige analytische Untersuchungen in linearen metrischen Ringen*, Japanese Journal of Mathematics, vol. 13(1936), pp. 61–80.
46. Hisami Nisigaki, *Über die durch hyperkomplexe Funktionen vermittelnden Abbildungen*, Tôhoku Mathematical Journal, vol. 44(1937), pp. 18–31.
47. B. J. Pettis, *Differentiation in Banach spaces*, this Journal, vol. 5(1939), pp. 254–269.
48. E. Picard, *Sur certains systèmes d'équations aux dérivées partielles généralisant les équations de la théorie des fonctions d'une variable complexe*, Journal de Mathématiques pures et appliquées, (4), vol. 8(1892), pp. 217–232.
49. F. Ringleb, *Beiträge zur Funktionentheorie in hyperkomplexen Systemen*, I, Rendiconti del Circolo Matematico di Palermo, vol. 57(1933), pp. 311–340.
50. G. Scheffers, *Verallgemeinerung der Grundlagen der gewöhnlich complexen Funktionen*, I, II, Berichte über die Verhandlungen der Sächsischen Akademie der Wissenschaften zu Leipzig, Mathematisch-physikalische Klasse, vol. 45(1893), pp. 828–848; vol. 46(1894), pp. 120–134.
51. Ludwig Schlesinger, *Neue Grundlagen für einen Infinitesimalkalkül der Matrizen*, Mathematische Zeitschrift, vol. 33(1931), pp. 33–61.
52. Hans Schwerdtfeger, *Beiträge zum Matrizen-Kalkül und zur Theorie der Gruppenmatrix*, Bonn, 1935 (a thesis).
53. Hans Schwerdtfeger, *Sur les fonctions de matrices*, Comptes Rendus hebdomadaires des Séances de l'Académie des Sciences, Paris, vol. 201(1935), pp. 414–416.
54. Hans Schwerdtfeger, *Über mehrdeutige Matrixfunktionen*, Compositio Mathematica, vol. 3(1936), pp. 380–390.
55. Hans Schwerdtfeger, *Les fonctions de matrices*, Actualités Scientifiques et Industrielles, no. 649, *Exposés de Géométrie*, Paris, 1938.
56. J. B. Shaw, *General vector calculus*, Transactions of the American Mathematical Society, vol. 24(1922), pp. 195–244.
57. Kenjiro Shoda, *Ein Kriterium für normale einfache hyperkomplexe Systeme*, Proceedings of the Imperial Academy of Japan, Tokyo, vol. 10(1934), pp. 195–197.
58. L. Sobrero, *Di una nova variabile ipercomplessa interessante la teoria dell'elasticità*, Atti della Reale Accademia Nazionale dei Lincei, Rendiconti, (6), vol. 19(1934), pp. 77–82.
59. L. Sobrero, *Alcuni teoremi della teoria delle funzioni ipercomplesse*, Atti della Reale Accademia Nazionale dei Lincei, Rendiconti, (6), vol. 19(1934), pp. 135–140.
60. L. Sobrero, *Algebra delle funzioni ipercomplesse e sue applicazioni alla teoria matematica dell'elasticità*, Reale Accademia d'Italia, Memorie della Classe di Scienze fisiche, matematiche e naturali, Roma, vol. 6(1934), pp. 1–64.
61. L. Sobrero, *Applicazione degli ipercomplessi ai problemi di elasticità piana*, Atti della Reale Accademia Nazionale dei Lincei, Rendiconti, (6), vol. 19(1934), pp. 479–483.

62. L. Sobrero, *Theorie der ebenen Elastizität unter Benutzung eines Systems hyperkomplexer Zahlen*, Hamburger Mathematische Einzelschriften, no. 17, Leipzig, 1934.

63. N. Spampinato, *Sulle funzioni di una variabile in un'algebra complessa ad n unità dotata di modulo* I, II, III, Rendiconti del Circolo Matematico di Palermo, vol. 57(1933), pp. 235–272; vol. 59(1935), pp. 185–227.

64. N. Spampinato, *Intorno ad un'algebra reale a 4 unità*, Atti della Reale Accademia Nazionale dei Lincei, Rendiconti, (6), vol. 19(1934), pp. 694–699.

65. N. Spampinato, *Estensione nel compo bicomplesso di due teoremi, del Levi-Civita e del Severi, per le funzioni olomorfe di due variabili complesse*, I, II, Atti della Reale Accademia Nazionale dei Lincei, Rendiconti, (6), vol. 22(1935), pp. 38–43, 96–102.

66. N. Spampinato, *Sulle funzioni totalmente derivabili in un'algebra reale o complessa dotata di modulo*, I, II, Atti della Reale Accademia Nazionale dei Lincei, Rendiconti, (6), vol. 21(1935), pp. 621–625, 683–687.

67. N. Spampinato, *Una proprietà caratteristica delle funzioni totalmente derivabili*, I, II, Atti della Reale Accademia Nazionale dei Lincei, Rendiconti, (6), vol. 21(1935), pp. 626–631, 688–692.

68. N. Spampinato, *Sulla Rappresentazione delle funzioni di variabile bicomplessa totalmente derivabili*, Annali di Matematica pura ed applicata, (4), vol. 14(1936), pp. 305–325.

69. N. Spampinato, *Sulle funzioni di variabili bicomplessa o biduale*, Scritti matematici offerto a Luigi Berzolari, published by Zanichelli, 1936, pp. 595–611.

70. N. Spampinato, *Sulle funzioni di una variabile in una delle quattro algebre complesse commutative del 3° ordine*, Atti della Reale Accademia Nazionale dei Lincei, Rendiconti, (6), vol. 26(1937), pp. 3–7.

71. Angus E. Taylor, *Analytic functions in general analysis*, Annali della Reale Scuola Normale Superiore di Pisa, (2), vol. 6(1937), pp. 277–292.

72. Angus E. Taylor, *On the properties of analytic functions in abstract spaces*, Mathematische Annalen, vol. 115(1938), pp. 466–484.

73. Angus E. Taylor, *Biharmonic functions in abstract spaces*, American Journal of Mathematics, vol. 60(1938), pp. 416–422.

74. S. Takahashi, *Zum verallgemeinerten Infinitesimalkalkül der Matrizen*, I, II, Japanese Journal of Mathematics, vol. 9(1932), pp. 27–46; vol. 9(1932), pp. 257–284.

75. H. W. Turnbull, *On differentiating a matrix*, Proceedings of the Edinburgh Mathematical Society, (2), vol. 1(1927), pp. 111–128.

76. H. W. Turnbull, *A matrix form of Taylor's theorem*, Proceedings of the Edinburgh Mathematical Society, (2), vol. 2(1930-31), pp. 33–54.

77. H. W. Turnbull, *Invariant theory of a general bilinear form*, Proceedings of the London Mathematical Society, (2), vol. 33(1932), pp. 1–21.

78. Robert Wagner, *Les fonctions multivalentes dans l'espace des matrices*, Comptes Rendus hebdomadaires des Séances de l'Académie des Sciences, Paris, vol. 205 (1937), pp. 210–212.

79. Édouard Weyr, *Note sur la théorie des quantités complexes formées avec n unités principales*, Bulletin des Sciences Mathématiques, (2), vol. 11(1887), pp. 205–215.

80. J. Williamson, *Bazin's matrix and other allied matrices*, Proceedings of the Edinburgh Mathematical Society, (2), vol. 2(1929), pp. 240–251.

81. G. Zwirner, *Sulla integrazione, secondo Volterra di una matrice*, Atti della Reale Accademia Nazionale dei Lincei, Rendiconti, (6), vol. 23(1936), pp. 723–727.

University of Wisconsin and Delta State Teachers College.

ORDERLY DIFFERENTIAL SYSTEMS

By Joseph Miller Thomas

1. **Introduction.** Riquier has given two general existence theorems for systems of partial differential equations in a field of holomorphic functions. The first of these concerns orthonomic systems [8, 254].[1] Because it is locally applicable without any restriction being placed on the holomorphic functions defining the equations and the initial determination, it can conveniently be called an unrestricted theorem. The second is a restricted theorem generalizing the first. It is applicable both to orthonomic systems and to some non-orthonomic systems whose defining functions and initial determinations are subjected to certain inequalities [8, 384, 387].

The present paper defines an orderly system (§27) as a system decomposable into a finite number of orthonomic systems and proves for such a system an unrestricted existence theorem (Theorem 29.1). The class of orderly systems thus includes the orthonomic as a proper subclass. It also contains a subclass for which Riquier's non-orthonomic theorem does not give even a restricted result.

Without seriously complicating the analysis the method used here can be employed to prove a restricted theorem for a class of systems including as a proper subclass all systems covered by Riquier's two theorems. In order not to complicate the ideas now to be presented, the discussion of this generalization will be postponed.

The proof of the theorem for orderly systems is believed to have advantages over those previously given for orthonomic systems [2], [8], [9, 135–156], [13]. For this reason it seems desirable to make the present treatment self-contained. The chief feature is the absence of the auxiliary integers called cotes by Riquier. Their elimination makes available a simple direct means for testing whether a given system is in orderly or in orthonomic form. The corresponding test for orthonomic systems defined in the old manner is known, but its application involves a somewhat extended knowledge of the theory of linear inequalities [10].

2. **Notation.** The unknowns and independent variables will be denoted respectively by u_α ($\alpha = 1, 2, \cdots, r$) and x_i ($i = 1, 2, \cdots, n$), where r, n are arbitrary fixed positive integers. For the derivative

$$\frac{\partial^{i_1+i_2+\cdots+i_n} u_\alpha}{\partial x_1^{i_1}\,\partial x_2^{i_2}\cdots\partial x_n^{i_n}} \tag{2.1}$$

Received June 1, 1940.

[1] References to the bibliography at the end of the paper are given in brackets. The first number designates the entry and subsequent numbers, not otherwise described, the pages.

we employ the notation

$$\frac{\partial u_\alpha}{\partial m}, \qquad m = x_1^{i_1} x_2^{i_2} \cdots x_n^{i_n}. \tag{2.2}$$

The degree of the monomial m in x_k thus denotes the number of differentiations with respect to x_k. The letter D will be used to denote a derivative, and subscripts will be placed upon the D when it is necessary to consider several different derivatives simultaneously.

The monomial

$$u_\alpha x_1^{i_1} x_2^{i_2} \cdots x_n^{i_n} \tag{2.3}$$

will be called *the monomial* of the derivative in (2.2). We shall write $D \sim m$ to indicate that m is the monomial of D. A monomial like the m in (2.2) will be called *pure* to distinguish it from monomials like (2.3).

The system S we shall consider comprises a finite number of equations of the form

$$D' = f(x, D''). \tag{2.4}$$

The symbol D' represents a partial derivative of some unknown. It will be called a *left derivative*. Both D' and the equation (2.4) of which it is the left derivative will be said to *belong* to the unknown u_α appearing in D'. The symbol f represents a given function. Both D' and f vary in general from one equation to the next.

For convenience, we imagine the equations of S tabulated so that the sign = appears exactly once on each row. Two equations are *distinct* if and only if they are on different rows. The *number of equations* is the number of rows. The k-th equation is that on the k-th row. Thus

$$\frac{\partial u}{\partial x} = xu,$$

$$\frac{\partial u}{\partial x} = xu$$

is regarded as a system of two equations, with the first equation distinct from the second. The two left derivatives, however, are not regarded as distinct.

At times it is desirable to denote the system by the notation

$$D'_i = f_i(x, D'') \qquad (i = 1, 2, \cdots, \kappa) \tag{2.5}$$

which assigns specific symbols for each left derivative and right member.

The argument x on the right of (2.4) indicates the possible presence of some or all of the independent variables. The D'' denotes the presence of a finite set of derivatives of the unknowns. This set varies in general from equation to equation. Each D'' will be called a *right derivative* for the equation or equations in which it occurs.

Either D' or D'' may equal an unknown, that is, the word "derivatives" is to be interpreted as including the unknowns as derivatives of order zero.

The following example will be used to illustrate most of the results in the rest of the paper. It involves two unknowns u, v, two independent variables x, y and four equations. It will be cited as system $\mathfrak{S}$.

$$\frac{\partial u}{\partial x} = \frac{\partial u}{\partial y} + u,$$

$$\frac{\partial^3 u}{\partial x^2 \partial y} = 2\frac{\partial^2 u}{\partial y^2} + \frac{\partial v}{\partial y},$$

$$\frac{\partial v}{\partial y} = \frac{\partial^3 u}{\partial y^3} + \frac{\partial u}{\partial y},$$

$$\frac{\partial v}{\partial y} = \frac{\partial^3 u}{\partial y^3} + \frac{\partial^2 u}{\partial x^2} - \frac{\partial^2 u}{\partial x \partial y} - u.$$

3. **Numerical determination.** The totality of arguments (x's and D's) appearing in a system S will be denoted by $y_1, y_2, \cdots, y_r$. A set of constants which substituted for the y's satisfy the equations of S will be called a *numerical determination* for S. Later (§9) we shall prove that S is equivalent to a system having numerical determination $y_1 = y_2 = \cdots = y_r = 0$, or as we shall say, *zero numerical determination.* System $\mathfrak{S}$ has, among others, a zero numerical determination.

Included in the numerical determination is a set of values x_{i0} for the independent variables. These are often called *initial values* for the x's.

4. **The first inequality system.** Let D' be the left derivative of an equation of S and let D'' be a right derivative in the *same* equation. Suppose $D' \sim m'$, $D'' \sim m''$ and write the inequality

$$m' > m''. \tag{4.1}$$

The set obtained by making all possible choices of D', D'', subject to the restriction given above, plus the set

$$x_i > 1 \qquad (i = 1, 2, \cdots, n), \tag{4.2}$$

is the *first inequality system* for S. It is *consistent* if it has a solution in which the x's and the u's are positive constants. There is a simple test for consistency [12].

The first inequality system for $\mathfrak{S}$ is

$$\begin{array}{cccc} & ux > uy, & ux > u, & \\ & ux^2y > uy^2, & ux^2y > vy, & \\ & vy > uy^3, & vy > uy, & \\ vy > uy^3, & vy > ux^2, & vy > uxy, & vy > u, \\ & x > 1, & y > 1. & \end{array}$$

To apply the test of [12, §3] first we divide both members of each inequality by their highest common factor. Then we multiply the inequalities with u on the left by every inequality having u on the right. We find as the only independent inequalities free of u, v

$$x > y > 1.$$

Hence the first inequality system is consistent. Often, as in §36, it is easiest to prove the inequality system consistent by guessing a numerical solution for it. Here a solution is $x = 3$, $y = 2$, $u = 1$, $v = 5$.

5. **Ordering by inequality system.** Suppose the first inequality system for S is consistent. Fix upon a definite one of its solutions. Evaluated for this solution the monomial m corresponding to each derivative D, whether D appears in S or not, becomes a positive number k, which we shall call the *ordinal* of D with respect to S.

We write $D_1 > D_2$ and say that D_1 *follows* D_2 if the corresponding ordinals k_1, k_2 are in the relation $k_1 > k_2$. In this way, the solution of the first inequality system establishes certain order relations among the derivatives of the unknowns. Of course, k_1 may equal k_2 even when D_1 is distinct from D_2, so that the derivatives may not be completely ordered by this device.

For the solution of the first inequality system specifically given in §4, the derivatives appearing in system $\mathfrak{S}$ have the ordinals indicated below. Since the ordinals form an increasing sequence, each derivative follows those to its left.

u	$\dfrac{\partial u}{\partial y}$	$\dfrac{\partial u}{\partial x}$	$\dfrac{\partial^2 u}{\partial y^2}$	$\dfrac{\partial^2 u}{\partial x \partial y}$	$\dfrac{\partial^3 u}{\partial y^3}$	$\dfrac{\partial^2 u}{\partial x^2}$	$\dfrac{\partial v}{\partial y}$	$\dfrac{\partial^3 u}{\partial x^2 \partial y}$
1	2	3	4	6	8	9	10	18

The *ordinal of an equation* is the ordinal of its left derivative.

6. **Properties of the ordering.** It is readily seen that the ordering just discussed has the following properties:[2]

(a) $D_1 > D_2$ and $D_2 > D_3$ imply $D_1 > D_3$;

(b) $D_1 > D_2$ implies $\partial D_1/\partial x > D_2$ for every x;

(c) $D_1 > D_2$ implies $\partial D_1/\partial x > \partial D_2/\partial x$ for every x.

If, for example, $D_1 \sim m_1$ and $D_2 \sim m_2$, the relations $m_1 > m_2$ and $x > 1$ imply $m_1 x > m_2$ so that if the first inequality system establishes the relation $D_1 > D_2$, it also establishes the relation $\partial D_1/\partial x > D_2$.

7. **The order matrix.** In this section the word "order" means the total number of differentiations in a derivative. Let the left derivative of the i-th equation of S have order ω_i. Let the right member of the i-th equation have formal order ω_{ij} in the unknown to which the j-th equation belongs; that is, if the left member of the j-th equation is a derivative of u_α, then ω_{ij} is the order of the highest derivative of u_α appearing as an argument in the right member of the i-th equation.

[2] These are part of the defining properties of the canonical ordering in [11, 8].

When we say that ω_{11} is one for the equation

$$\frac{\partial u}{\partial x} = v\frac{\partial u}{\partial y},$$

there is no implication that the function v is different from zero. Moreover, if no derivative of u_α appears in a given function, the formal order of that function in u_α is defined to be $-\infty$.

It is convenient to denote by α_i the index of the unknown to which the i-th equation belongs. From the definition of the ω's we note that

$$\omega_{ij} = \omega_{ik} \qquad (\alpha_j = \alpha_k). \tag{7.1}$$

We shall write the ω's as a matrix whose i-th row and first column contain ω_i and whose i-th row and $(j+1)$-th column contain ω_{ij}.

The order matrix for system $\mathfrak{S}$ is

$$\begin{array}{c|cccc} 1 & 1 & 1 & -\infty & -\infty \\ 3 & 2 & 2 & 1 & 1 \\ 1 & 3 & 3 & -\infty & -\infty \\ 1 & 3 & 3 & -\infty & -\infty \end{array}.$$

8. **Defining properties.** The systems to be studied in the rest of this paper all satisfy the five following conditions.

(i) The equations have the form

$$D' = f(x, D'').$$

(ii) The unknowns, the independent variables, the derivatives and the equations appearing explicitly are finite in number.

(iii) A numerical determination (§3) is given.

(iv) The functions f are holomorphic[3] about the numerical determination.

(v) The first inequality system (§4) is consistent.

The derivatives will be considered to be ordered by the first inequality system.

Such a system will be called *admissible*. The numerical determination will be conceived as forming an integral part of the system. Since system $\mathfrak{S}$ is satisfied by making all its arguments zero, we are justified in attaching to it the zero numerical determination. Suppose this done. Then it is readily seen that system $\mathfrak{S}$ is admissible.

9. **Transformation of arguments.** Put

$$y_\tau = z_\tau + w_\tau \qquad (\tau = 1, 2, \cdots, \nu), \tag{9.1}$$

where the y's are the arguments of S, the z's are new arguments, and the w's are functions of the x's, holomorphic about the initial values and reducing to

[3] We employ "holomorphic" as equivalent to "analytic" in [6, 7].

the numerical determination of S when the x's are given their initial values. The substitution (9.1) converts S into a system T in the arguments z. Evaluation of (9.1) for the numerical determination of S gives

$$z_r = 0. \tag{9.2}$$

If T is given the numerical determination (9.2), the right members of T are holomorphic in the z's about the numerical determination for T since they equal functions holomorphic in the y's about the numerical determination for S, see [6, 51, Theorem 1].

Among the y's are included the x's. In the present section we find it convenient to assume that the equations (9.1) transforming the x's have the form

$$x_i = x_i' + x_{i0} \qquad (i = 1, 2, \cdots, n), \tag{9.3}$$

where x_{i0} are the initial values of the x's, that is, we assume that the w's corresponding to the x's are constant. As a consequence of (9.3) we have

$$\frac{\partial}{\partial x_i} = \frac{\partial}{\partial x_i'}. \tag{9.4}$$

Given a transformation

$$u_\alpha - \bar{u}_\alpha = v_\alpha \qquad (\alpha = 1, 2, \cdots, r), \tag{9.5}$$

where the v's are new unknowns and the $\bar{u}$'s are functions of the x's holomorphic about the initial values, by adjoining (9.3) and appropriate equations obtained by applying (9.4) to (9.5), we can extend (9.5) to give a transformation (9.1) on *all* the arguments of S. By the transformation so extended S is converted into a system T in the unknowns v with right members which are holomorphic about the zero numerical determination [6, 51, Theorem 1].

Moreover, if the derivatives y_r, z_r correspond respectively to the monomials m_r, n_r, then n_r arises from m_r by replacing the letters u, x by v, x'. The first inequality system for S is the same as that for T except for the letters used to denote the unknowns. Accordingly, the transformation (9.1) extending (9.5) converts S into a system T having the same first inequality system as S.

Hence it is seen that T is admissible if S is.

10. **Evaluation on a variety.** A given subset of the independent variables will be denoted by the monomial m which is the product of those variables. The equation $m = 0$ will mean that all the variables of m are to be made zero in the expression concerned. In this connection, the monomial 1 will mean a vacuous set of the x's, and when m becomes unity the equation $m = 0$ will mean to make no x's zero.

Let $F(x, D)$ be a function holomorphic about a numerical determination N. Let there be given corresponding to each u a function φ holomorphic in the x's about the initial values included in N. Put

$$u = \varphi(x). \tag{10.1}$$

If the φ's when differentiated and evaluated for the initial values give N, it is possible to substitute the φ's for the u's, and the results $\psi(x)$ of the substitution are holomorphic about the initial values [6, 51, Theorem 1]. The function ψ will be called the *value of F on the variety* (10.1) or on the variety φ.

A variety φ gives a *solution* of the equation $F = 0$ if $\psi = 0$.

The function $\psi(x)$ further evaluated for $m = 0$ will be called the *value of F on the variety* φ_m. Because of [6, 51, Theorem 1], to evaluate F on φ_m we may evaluate the extended form of (10.1) for $m = 0$ and substitute the results in F.

11. **Principal and parametric derivatives.** If $D_1 \sim m_1$ is a derivative of $D_2 \sim m_2$, then $m_1 = \lambda m_2$, where λ is a pure monomial. This relation will be expressed by saying that "m_1 is a multiple of m_2", it being understood that only x's appear in the ratio λ.

The left members of S and all their derivatives are called *principal* for S. A derivative of an unknown is called *parametric* for S if it is the derivative of no left member of S. These names are also applied to the corresponding monomials.

An unknown is *principal* if at least one of its derivatives is a left derivative; otherwise it is *parametric*. A principal unknown is, of course, not necessarily a principal derivative. It is occasionally convenient to suppose the notation has been adjusted so that the principal unknowns are

$$u_\alpha \qquad (\alpha = 1, 2, \cdots, l). \tag{11.1}$$

12. **Prolonged systems.** If $y_1, y_2, \cdots, y_\nu$ are differentiable functions of $x_1, x_2, \cdots, x_n$, the formulas

$$\delta_i = \frac{\partial y_1}{\partial x_i}\frac{\partial}{\partial y_1} + \frac{\partial y_2}{\partial x_i}\frac{\partial}{\partial y_2} + \cdots + \frac{\partial y_\nu}{\partial x_i}\frac{\partial}{\partial y_\nu} \tag{12.1}$$

define a set of differential operators applicable to any differentiable function $F(y_1, y_2, \cdots, y_\nu)$.

Since holomorphic functions are by definition differentiable, the set (12.1) can be applied to each equation of a system S. The result is another equation of the form (2.4). Let all such equations, arising from a single application of (12.1) to equations of S, be adjoined to S. The resulting system S' is called the *(first) prolonged system of* S. It contains $n + 1$ times as many equations as does S.

Suppose the operator δ_1 is applied to equation (2.4). If D' has monomial m', the left derivative of the new equation has monomial $m'x_1$. The presence of D'' in (2.4) implies the presence of both $\partial D''/\partial x_1$ and D'' in the right member of the new equation; and only right derivatives of these two categories are present in the new equation. The first inequality system for S' can be obtained by adjoining to that for S inequalities of the types

$$m'x_1 > m''x_1, \qquad m'x_1 > m'', \tag{12.2}$$

where

$$m' > m'' \tag{12.3}$$

belongs to the first inequality system for S. In the presence of $x_1 > 1$, (12.3) implies both inequalities (12.2). Accordingly, the first inequality system of S' is consistent if that of S is. Hence S' is admissible as far as (i), (ii), (v) are concerned. The proof that S' is admissible will be completed in §§17–18. It will also be proved that every holomorphic solution of S satisfies S'.

The k-th prolonged system S^k is defined by induction. The infinite set obtained by letting k increase without limit is denoted by S^∞ and is called the *prolongation*.

13. **The monomial sets.** Consider the totality of left derivatives in a system S which involve a particular unknown u. Denote by M' the set obtained by removing the u from the corresponding monomials. Let the least common multiple of the monomials in M' be

$$x_1^{j_1} x_2^{j_2} \cdots x_n^{j_n}. \tag{13.1}$$

The finite set M consisting of all distinct monomials which divide (13.1) and have coefficient equal to one is called the *monomial set* for the corresponding u with respect to S. The set M thus consists of all monomials whose exponents $k_1, k_2, \cdots, k_n$ satisfy

$$j_1 \geqq k_1 \geqq 0, \cdots, j_n \geqq k_n \geqq 0. \tag{13.2}$$

The set M falls into two mutually exclusive parts. A monomial of M belongs to the *principal set* M^* if it is divisible by at least one monomial of M'. The set M^* contains the distinct monomials of M' and in particular cases coincides with it. A monomial of M belongs to the *parametric set* $\bar{M}$ if it is divisible by no monomial of M'.

It should be noted that two or more monomials in M' may coincide, whereas such coincidences are impossible in the sets M, M^*, $\bar{M}$.

If the monomial

$$m = x_1^{k_1} x_2^{k_2} \cdots x_n^{k_n} \tag{13.3}$$

of the set M has degree in x_i equal to the maximum possible, that is, if k_i is equal to the corresponding j_i in (13.1), then x_i is a *multiplier* for m; otherwise x_i is a *non-multiplier* for m. In particular, the least common multiple (13.1) has all the x's for multipliers and no non-multipliers.

We shall denote by x_* a multiplier of m, by x_0 a non-multiplier, by m^* the product of the multipliers and by m^0 the product of the non-multipliers.

In system $\mathbb{S}$, there are two sets M', one M'_1 for the left derivatives of u and one M'_2 for those of v.

M'_1: $\quad x, \quad x^2y$

M'_2: $\quad y, \quad y$

The corresponding monomial sets are given below.

M_1^* :	$x,\ x^2,\ xy,\ x^2y$	$\overline{M}_1$:	$1,\ y$
M_2^* :	y	$\overline{M}_2$:	1

We find it convenient to denote the monomials by their exponents and to use 11–01 to mean "the monomial xy with non-multiplier x and multiplier y". In this compact notation the two sets M_1, M_2 for $\mathfrak{S}$ are as given below.

M_1 :	**00–00** ***10–00** 20–10 01–01 11–01 *21–11	M_2 :	00–10 *01–11

It is convenient to write the monomials in order of increasing rank (cf. §34). The asterisk is used to denote the monomials in the original set M'.

If the monomial m given by (13.3) is in M, its exponents satisfy (13.2). If, moreover, x_i is non-multiplier for m, then $j_i > k_i$ so that $j_i + 1 \geqq k_i$. Hence the exponents of the monomial mx_i also satisfy (13.2) and consequently mx_i also belongs to M. We have therefore a well-known result (cf. [11, 57])

THEOREM 13.1. *Every product of a monomial of M by a non-multiplier belongs to M. Every product of a monomial of the principal set M^* by a non-multiplier belongs to M^*.*

On the other hand, the product of a monomial of $\bar{M}$ by a non-multiplier may belong either to $\bar{M}$ or to M^*.

The set M always contains the monomial 1. Hence an arbitrary monomial p in $x_1, x_2, \cdots, x_n$ is divisible by at least one monomial of M. Among all the monomials in M dividing p let m be one having maximum degree. Suppose the quotient p/m involves an x_i which is a non-multiplier for m. By Theorem 13.1 mx_i belongs to M. As it divides p and has higher degree than m, there is a contradiction. Hence p/m involves only multipliers of m.

Moreover, there is only one m which is in M, which has maximum degree and which divides p. For if there are two, say m, m', we have

$$p = mq = m'q'.$$

Suppose the exponent of x_1 is greater in m than in m'. Then x_1 is a divisor of q' and therefore a multiplier of m'. This is impossible, for the exponent of x_1 in m' is less than in m and therefore less than the maximum in M. Hence the exponent of x_1 is the same in m and m'. As the same argument applies to all the x's, we conclude $m = m'$. The unique monomial m is called the *generator* of all of its multiples by multipliers. We have proved the partly known result (cf. [11, 57])

THEOREM 13.2. *With an arbitrary pure monomial p there is associated in M a unique monomial m called its generator and characterized by being the monomial of*

highest degree in M which divides p. The monomial p is the product of its generator by multipliers of the generator.

If M' is vacuous, there is no least common multiple. Hence M^*, $\bar{M}$ are undefined. We shall, however, interpret M^* as vacuous and $\bar{M}$ as consisting of the monomial 1 with all the independent variables for multipliers.

14. **The Maclaurin expansion.** We shall seek to determine for the unknown u's infinite series which converge about $x_1 = \cdots = x_n = 0$ and which satisfy a given admissible system S. Because of the result in §9, the particular choice of initial values does not restrict the generality of the results to be obtained.

Accordingly, we assume an expansion for each u in the form

$$u = \sum_0^\infty a_{i_1 i_2 \cdots i_n} x_1^{i_1} x_2^{i_2} \cdots x_n^{i_n} \tag{14.1}$$

and seek to determine the a's. The results of the last section enable us to rewrite (14.1) in a useful form as a finite sum. To do this, we collect together all terms containing monomials having the same generator $x_1^{i_1}x_2^{i_2} \cdots x_n^{i_n}$, factor out the generator from their sum, and denote by $b_{i_1 i_2 \cdots i_n}$ the power series multiplying the generator. (In special cases, $b_{i_1 i_2 \cdots i_n}$ reduces to a constant.) This rearrangement gives

$$\sum x_1^{i_1} x_2^{i_2} \cdots x_n^{i_n} b_{i_1 i_2 \cdots i_n}, \tag{14.2}$$

where the summation extends over the monomials $x_1^{i_1}x_2^{i_2} \cdots x_n^{i_n}$ of M, and where each $b_{i_1 i_2 \cdots i_n}$ is a function of the multipliers of the monomial multiplying it in (14.2). If u is holomorphic, the b's are also holomorphic and u is equal to (14.2), see [6, 50, Theorem 1].

By means of (14.2), equation (14.1) can be rewritten as

$$u = \bar{u} + u^*, \tag{14.3}$$

where the monomials in the infinite series $\bar{u}$, u^* have generators in $\bar{M}$, M^* respectively. The functions $\bar{u}$, u^* will be called respectively the *parametric part* and the *principal part* [8, 170] of u. The names parametric and principal will also be applied to the coefficients. If u is a parametric unknown, M^* is vacuous and u consists entirely of its parametric part $\bar{u}$.

The finite sums (14.2) for u, v of system $\mathfrak{S}$ are

$$u = 1b_{00} + yb_{01}(y) + xb_{10} + x^2b_{20}(x) + xyb_{11}(y) + x^2yb_{21}(x, y),$$

$$v = 1c_{00}(x) + yc_{01}(x, y),$$

so that

$$\bar{u} = 1b_{00} + yb_{01}(y),$$

$$u^* = xb_{10} + x^2b_{20}(x) + xyb_{11}(y) + x^2yb_{21}(x, y),$$

$$\bar{v} = 1c_{00}(x),$$

$$v^* = yc_{01}(x, y).$$

The arguments of the b's and c's are the corresponding multipliers. We note in particular that b_{00}, b_{10} reduce to constants because the corresponding monomials have no multipliers.

If m, m' are monomials and m does not divide m', then (see §2 for the notation)

$$\frac{\partial m'}{\partial m} = 0 \qquad (m \nmid m').$$

Consequently, if m belongs to M^*, we have from (14.3)

$$\frac{\partial \bar{u}}{\partial m} = 0, \qquad \frac{\partial u}{\partial m} = \frac{\partial u^*}{\partial m}, \qquad (m \in M^*). \tag{14.4}$$

Putting

$$m = x_1^{i_1} x_2^{i_2} \cdots x_n^{i_n},$$

and letting as in §13 m^0 denote the non-multipliers of m, we have

$$\left(\frac{\partial u}{\partial m}\right)_{m^0=0} = \left[\frac{\partial}{\partial m}\,(x_1^{i_1} x_2^{i_2} \cdots x_n^{i_n} b_{i_1 i_2 \cdots i_n})\right]_{m^0=0}$$

because every term in the expansion of u not divisible by m disappears on differentiation and every multiple of m by non-multipliers vanishes when the non-multipliers are made zero. This enables us to compute the values of

$$\left(\frac{\partial u}{\partial m}\right)_{m^0=0} \tag{14.5}$$

when the b corresponding to m is given. Conversely, if (14.5) is given as a function F of the variables m^* and if $i_1 > 0$, integrate with respect to x_1 between the limits 0 and x_1. There results a function divisible by x_1. Application of this process i_1 times with respect to x_1, i_2 times with respect to $x_2, \cdots, i_n$ times with respect to x_n gives a function G divisible by m. The quotient G/m involves only the variables present in F and is equal to $b_{i_1 i_2 \cdots i_n}$.

15. **The initial determination.** For the expansion (14.1) we have

$$a_{i_1 i_2 \cdots i_n} = \frac{1}{i_1! i_2! \cdots i_n!}\left(\frac{\partial^{i_1+i_2+\cdots+i_n} u}{\partial x_1^{i_1} \partial x_2^{i_2} \cdots \partial x_n^{i_n}}\right)_0, \tag{15.1}$$

where the subscript 0 means evaluation for the initial values $x_1 = x_2 = \cdots = x_n = 0$. Hence division by an integer enables us to pass from the value of a derivative at the origin to the corresponding coefficient in u. In particular, the numerical determination consists essentially of certain coefficients of the u's together with the initial values of the x's.

Corresponding to each unknown u_α and to each monomial m in its parametric set $\bar{M}_\alpha$ assign a function $I_{\alpha m}$ of the multipliers subject to the two restrictions:

(1) $I_{\alpha m}$ is holomorphic in its arguments about the initial values;

(2) if the coefficient a of any monomial p in u_α belongs to the numerical determination for S, then the coefficient of p in $I_{\alpha m}$ has the same value a.

A set of functions I containing one $I_{\alpha m}$ for each combination α, m, where m belongs to $\bar{M}_\alpha$, is called a *relevant initial determination* for S. The system S, augmented by the equations

$$\left(\frac{\partial u_\alpha}{\partial m}\right)_{m^0=0} = I_{\alpha m} \qquad (m \in \bar{M}_\alpha), \tag{15.2}$$

is called a *determined system*.

We note that the nature of the initial determination, that is, the number of functions and their sets of arguments, is governed solely by the left derivatives of S.

From the discussion at the end of §14, applied to a monomial m of $\bar{M}_\alpha$, it is clear that imposing conditions (15.2) is equivalent to giving the b's in (14.2) or to giving the values of all the parametric coefficients.

The initial determination for system $\mathfrak{S}$ contains the values

$$u(0, 0), \quad \left(\frac{\partial u}{\partial y}\right)_{x=0}, \quad v(x, 0). \tag{15.3}$$

Let us convert $\mathfrak{S}$ into a determined system by assigning an initial determination. To conform with (2) we must place

$$u(0, 0) = \left(\frac{\partial u}{\partial y}\right)_{x=y=0} = \left(\frac{\partial^2 u}{\partial y^2}\right)_{x=y=0} = \left(\frac{\partial^3 u}{\partial y^3}\right)_{x=y=0} = 0.$$

Hence specifying the initial determination consists in giving to (15.3) values

$$0, \quad \varphi(y), \quad \psi(x), \tag{15.4}$$

where φ is an arbitrarily chosen function of y which is holomorphic about $y = 0$ and which together with its first two derivatives vanishes for $y = 0$ and ψ is an arbitrarily chosen function of x holomorphic about $x = 0$.

If in particular we choose

$$0, \quad 4y^3, \quad x, \tag{15.5}$$

we readily find

$$b_{00} = 0, \qquad yb_{01}(y) = \int_0^y 4t^3\,dt = y^4, \qquad c_{00}(x) = x$$

so that

$$\bar{u} = y^4, \qquad \bar{v} = x.$$

When I has once been assigned, the parametric parts $\bar{u}_\alpha$ are determined. Let the transformation (9.5) for such $\bar{u}$'s be applied to S. From (9.5) and (14.4) we deduce

$$\left(\frac{\partial v_\alpha}{\partial m}\right)_{m^0=0} = 0 \qquad (m \in \bar{M}_\alpha).$$

Comparison with (15.2) shows that the *initial determination for the system T in the v's is zero.*

Consider the determined system

$$\frac{\partial v}{\partial x} = \frac{\partial^2 v}{\partial y^2}, \qquad v(0, y) = \frac{1}{1 - y} \tag{15.6}$$

which has been given numerical determination

$$v(0, 0) = 1, \qquad \left(\frac{\partial v}{\partial x}\right)_{x=y=0} = \left(\frac{\partial^2 v}{\partial y^2}\right)_{x=y=0} = 2.$$

The change of unknown

$$v = u + \frac{1}{1 - y} + 2x$$

gives the equation

$$\frac{\partial u}{\partial x} = \frac{\partial^2 u}{\partial y^2} + \frac{2}{(1 - y)^3} - 2 \tag{15.7}$$

with numerical and initial determinations consisting of zeros.

16. **Determination of the coefficients.** In the last section, we saw that the numerical determination gives the values of some of the coefficients and in particular of some of the parametric coefficients. The other conditions imposed by the system on the parametric derivatives are in general not readily expressible (see §33) in the form (2.4). Consequently, we begin by ignoring any such conditions implied by S. From this viewpoint it is natural to seek a solution having an arbitrary initial determination, that is, one in which the parametric b's are holomorphic functions chosen arbitrarily except for the coefficients in the numerical determination.

17. **The principal system.** The parametric coefficients being supposed given, our next task is to specify a process which determines one and only one value for each principal coefficient.

If m is a monomial in the principal set M_α^* for u_α, there is at least one equation in the prolongation whose left derivative corresponds to $u_\alpha m$. Of all such equations, make an arbitrary choice of exactly one for each monomial in the sets M_α^* to form a system A.

Separate the equations of A into sets of equal ordinal

$$A_1, A_2, \cdots, A_e$$

so that every left derivative in A_i follows every left derivative in A_j if $i > j$, whereas no left derivative in A_i follows a left derivative in the same set A_i.

The equations of A_1 all belong to the original system S. Hence a numerical determination for A_1 is given by the numerical determination for S. Moreover, every right derivative in A_1 is parametric: if it were not, it would have to follow some left derivative in A_1. Since only parametric derivatives appear on the right of equations in A_1, system A_1 is said to be in *solved form*.

The derivatives which appear as arguments of A but not of S enter linearly. Hence the functions defining A are holomorphic about any set of values

built up from the numerical determination for S by assigning arbitrary values to the additional arguments.

A right derivative in A_2, if principal, must be a left derivative in A_1. If the series for the principal right derivatives in A_2 given by A_1 are substituted in the right members of A_2, there result functions holomorphic about the numerical values indicated above [6, 51, Theorem 1]. In this way, A_2 is also put in solved form. From the solved form values to complete the numerical determination are obtained by getting the parametric coefficients from the initial determination and then computing the principal by substitution.

A continuation of this process shows that A can be put in solved form S^*. It is readily seen that S^* is admissible. We shall call S^* a *principal system* for S.

For system $\mathfrak{S}$ a principal system $\mathfrak{S}^*$ is, multipliers and non-multipliers being indicated to the right,

$$
\mathfrak{S}^* \qquad
\begin{array}{rcl@{\qquad}c@{\qquad}c}
 & & & m^* & m^c \\
\dfrac{\partial u}{\partial x} &=& \dfrac{\partial u}{\partial y} + u, & 1 & xy \\[2ex]
\dfrac{\partial^2 u}{\partial x^2} &=& \dfrac{\partial^2 u}{\partial y^2} + 2\dfrac{\partial u}{\partial y} + u, & x & y \\[2ex]
\dfrac{\partial^2 u}{\partial x \partial y} &=& \dfrac{\partial^2 u}{\partial y^2} + \dfrac{\partial u}{\partial y}, & y & x \\[2ex]
\dfrac{\partial^3 u}{\partial x^2 \partial y} &=& \dfrac{\partial^3 u}{\partial y^3} + 2\dfrac{\partial^2 u}{\partial y^2} + \dfrac{\partial u}{\partial y}, & xy & 1 \\[2ex]
\dfrac{\partial v}{\partial y} &=& \dfrac{\partial^3 u}{\partial y^3} + \dfrac{\partial u}{\partial y}. & xy & 1
\end{array}
$$

Note that the principal derivatives have been eliminated from the right members and that the last equation of $\mathfrak{S}$ has arbitrarily been selected for omission in preference to the third.

In §10 we saw that the result of evaluating

$$D' - f \tag{17.1}$$

for u's holomorphic about the numerical determination is a holomorphic function F. The derivatives of F can be found by applying (12.1) to (17.1), see [6, 8]. If the u's constitute a solution of S, (17.1) is zero and so is its derivative. Consequently, any solution of S satisfies A. Now in deducing S^* from A we replace certain arguments y_λ by others z_λ. For any solution of S y_λ and the corresponding z_λ reduce to exactly the same holomorphic function of the x's. Hence every solution of S is a solution of S^*.

Note that there are numerous principal systems for a given system S, but our process calls for the arbitrary choice of one of them.

18. **The tentative solution.** The left derivatives of S^* are in one-to-one correspondence with the monomials of the principal sets M^*_α. To each equa-

tion of S^* are assigned the multipliers and the non-multipliers of the corresponding monomials in the sets M_α^*. Form the first derivative of each equation in S^* with respect to each of the corresponding multipliers. Just as in the case of the principal system (§17) the resulting system can be put in solved form. The solved form S_1 is the *(first) extended system of S with respect to S^**. Note that S_1 contains the solved form of some of the equations in S' but none of the equations in S^*. The k-th extended system is defined by induction. In this connection, the multipliers for an equation in S_{k-1} are to be interpreted as the multipliers of the equation in S^* from which the equation in S_{k-1} arose. The totality of all extended systems determined by S^* is called the *extension of S with respect to S^**.

Since each principal monomial can be generated in one and only one way by multiplying the principal set by multipliers, each principal derivative occurs once and only once as a left derivative in the extension.

If we put $x_1 = x_2 = \cdots = x_n = 0$ and the corresponding values of the parametric derivatives in the extension, we determine values for all the principal coefficients. Consequently, *given any admissible system and a relevant initial determination, we can construct a unique expansion for each unknown u_α*. This set of r expansions will be called a *tentative solution.*

It is convenient to state

THEOREM 18.1. *A determined admissible system cannot have more than one solution.*

19. **The derived systems.** Next we must discuss the convergence of the tentative solution. For this purpose it is desirable to differentiate the equations of S^* so as to obtain a system in which all left derivatives of each principal u_α have the same order, which of course has to be at least equal to the corresponding g_α, where g_α is the degree of the least common multiple of the monomials in M_α^*.

Let h_α be positive integers satisfying $h_\alpha \geqq g_\alpha$. Consider the principal part of the finite sum (14.2) for the unknown u_α. Let the degree of one of the monomials m multiplying a b be $h_\alpha - s$. The corresponding b can be written

$$b = \sum jc + \sum kd, \tag{19.1}$$

where j ranges over all monomials of degree less than s in the multipliers of m, where the c's are constants, where k ranges over all monomials of degree s in the multipliers of m, and where the d's are holomorphic functions in the multipliers of m.

If we agree to compute the principal coefficients c and then regard them as part of the initial determination, the equation in S^* with left derivative corresponding to mu_α can be replaced by equations with left derivatives corresponding to kmu_α. These equations are formed by applying the differential operator $\partial/\partial k$ to the equation of S^* with left member corresponding to mu_α. All of the left derivatives so obtained have order $(h_\alpha - s) + s = h_\alpha$.

Application of this process for all m's and for all u's gives a system with all left derivatives of u_α of order h_α. The equations of this system can be put in solved form $S_{h_1 h_2 \cdots h_l}$ as was the principal system in §17. The system is admissible. Moreover, if S has consistent first inequality system, so also has $S_{h_1 h_2 \cdots h_l}$. The system $S_{h_1 h_2 \cdots h_l}$ will be called the *derived system of index* $h_1 h_2 \cdots h_l$, the principal unknowns being $u_1, u_2, \cdots, u_l$.

If for $\mathfrak{S}^*$ we take $h_1 = 3$, $h_2 = 1$, we find $\mathfrak{S}_{31}$ contains

$$\begin{array}{lcc}
 & m^* & m^0 \\
\dfrac{\partial^3 u}{\partial x^3} = \dfrac{\partial^3 u}{\partial y^3} + 3\dfrac{\partial^2 u}{\partial y^2} + 3\dfrac{\partial u}{\partial y} + u, & x & y \\
\dfrac{\partial^3 u}{\partial x \partial y^2} = \dfrac{\partial^3 u}{\partial y^3} + \dfrac{\partial^2 u}{\partial y^2}, & y & x \\
\dfrac{\partial^3 u}{\partial x^2 \partial y} = \dfrac{\partial^3 u}{\partial y^3} + 2\dfrac{\partial^2 u}{\partial y^2} + \dfrac{\partial u}{\partial y}, & xy & 1 \\
\dfrac{\partial v}{\partial y} = \dfrac{\partial^3 u}{\partial y^3} + \dfrac{\partial u}{\partial y}. & xy & 1
\end{array}$$

The initial determination contains, in addition to (15.3), the values

$$\left(\frac{\partial u}{\partial x}\right)_{x=y=0} = \left(\frac{\partial u}{\partial y}\right)_{x=y=0} + u(0, 0) = 0,$$

$$\left(\frac{\partial^2 u}{\partial x^2}\right)_{x=y=0} = \left(\frac{\partial^2 u}{\partial y^2}\right)_{x=y=0} + 2\left(\frac{\partial u}{\partial y}\right)_{x=y=0} + u(0, 0) = 0,$$

$$\left(\frac{\partial^2 u}{\partial x \partial y}\right)_{x=y=0} = \left(\frac{\partial^2 u}{\partial y^2}\right)_{x=y=0} + \left(\frac{\partial u}{\partial y}\right)_{x=y=0} = 0.$$

The transformation (§15)

$$u' = u - \int_0^y \varphi(t)\, dt,$$

$$v' = v - \psi(x),$$

carries $\mathfrak{S}_{31}$ into a system in the unknowns u', v' with zero initial determination.

20. The second inequality system. Let the equations of S^* be numbered by the indices $1, 2, \cdots, \kappa^*$. Denote the elements of the order matrix for S^* by ω^*'s. The system

$$\omega_i^* + t_i \geqq \omega_{ji}^* + t_j \qquad (i, j = 1, 2, \cdots, \kappa^*) \tag{20.1}$$

is called the *second inequality system* for S^*. The consistency conditions for it are

$$\sum (\omega_i^* - \omega_{ji}^*) \geqq 0, \tag{20.2}$$

where i is summed over an arbitrary cycle of arbitrary length from $(1, 2, \cdots, \kappa^*)$ and j has the value immediately following that of i in the cycle [12, §7]. Suppose conditions (20.2) satisfied.

Since (20.1) has a solution t_i, it also has the solution $t_i + a$, where a is an arbitrary positive number. Hence if (20.1) is consistent, we can choose positive integers t_i satisfying both (20.1) and

$$\omega_i^* + t_i \geqq g_{\alpha_i} + 1, \tag{20.3}$$

where α_i is the index of the unknown to which the i-th equation of S^* belongs. Suppose such a choice made.

Now form the derived system S^{**} whose index is given by the left members of (20.3). To do this, the i-th equation of S^* is differentiated t_i times. Relations (20.1) say that the order h_α of the left derivatives belonging to u_α is the order of the resulting system in u_α. This property is not disturbed by putting the system in the solved form S^{**}.

The *second inequality system* for S is defined to be

$$\omega_i + t_i \geqq \omega_{ji} + t_j \qquad (i, j = 1, 2, \cdots, \kappa), \tag{20.4}$$

whose conditions of consistency are

$$\sum (\omega_i - \omega_{ji}) \geqq 0. \tag{20.5}$$

The consistency of the second inequality system for S implies that of the system for S^*.

21. Definition of orthonomic system. As we shall see in §37 the tentative solution does not always converge. In order to proceed we find it convenient to assume that S^{**} exists.

A system is *orthonomic* if it has the following properties.

(i) The equations have the form

$$D' = f(x, D'').$$

(ii) The unknowns, the independent variables, the derivatives and the equations appearing explicitly are finite in number.

(iii) A numerical determination (§3) is given.

(iv) The functions f are holomorphic about the numerical determination.

(v) The first inequality system (§4) is consistent.

(vi) The second inequality system (20.4) is consistent.

The system $\mathfrak{S}$ is readily seen to be orthonomic.

22. Dominant systems. Next we examine the tentative solution for convergence. We consider the derived system S^{**}.

The derivatives of highest order h_α in S^{**} enter linearly because they appear only as a result of the application of formulas (12.1). Consequently, we assume S^{**} in the form

$$D' = f, \tag{22.1}$$

where

$$f = \sum pD'' + q, \tag{22.2}$$

where D', D'' represent derivatives of u_α of order h_α and where the coefficients p, q involve only derivatives of u_α of order less than h_α. When the process of calculating the coefficients is applied to (22.1), it is found that every principal derivative is a polynomial with non-negative integral coefficients in the parametric derivatives and the partial derivatives of the f's. Hence

$$a = \pi(b, c), \tag{22.3}$$

where a, b, c represent the coefficients of u, of the initial determination and of f respectively, and where π is a polynomial with non-negative integral coefficients. Consequently, the principal coefficients of any solution of a system like (22.1) are non-negative provided the coefficients of the f's and the parametric coefficients are non-negative.

We suppose without loss of generality (§15) that the initial determination of S^{**} is zero. Every principal coefficient is then a polynomial with non-negative integral coefficients in the coefficients of the f's. Moreover, the parametric unknowns are all replaced by zero, so that all the unknowns are to be regarded as principal.

Consider a particular equation (22.1) with left derivative $D' \sim u_\alpha m'$. Corresponding to it write an equation

$$D_1 = F(x, D, E_1, E_2) \tag{22.4}$$

in the unknowns U, where $D_1 \sim U_\alpha m'$. The argument D on the right of (22.4) means the presence of every $D \sim U_\beta m''$, where $u_\beta m'' \sim$ a derivative (necessarily parametric) of order less than h_β (§20) in system (22.1). The argument E_1 likewise denotes the presence of every derivative $E_1 \sim U_\beta m''$, where $u_\beta m'' \sim$ a parametric or principal derivative of order h_β which precedes the left derivative of the corresponding equation (22.1). Similarly, $E_2 \sim U_\beta m''$, where $u_\beta m'' \sim$ a parametric or a principal derivative which does not precede the left member of the corresponding equation (22.1). The set $E_1 + E_2$ therefore contains all derivatives of each unknown u_γ of the corresponding order h_γ.

Treating all equations of (22.1) in this way we find a system (22.4) whose equations and left members are in one-to-one correspondence with (22.1).

Note that the sets D, $E_1 + E_2$ are the same for all right members, whereas the separation of the derivatives of order h_β into E_1, E_2 may vary from equation to equation.

Denote by H_{E_1} the number of E_1's which appear in a given right member and which belong to the same unknown as the E_1 appearing as the subscript on H. The number H_{E_1} varies from equation to equation. An integer H_{E_2} is similarly defined.

The arguments in a given equation (22.1) correspond to a (proper or improper) subset of the x, D, E_1 in the corresponding equation (22.4). Because

the sets D, E_2 in general contain derivatives following the left derivative, system (22.4) is not admissible. Suppose, however, that each argument in (22.4) which does not have a correspondent in (22.1) is multiplied by a constant parameter λ. Then by applying to (22.4) the operations applied to (22.1) in finding the coefficients of its tentative solution we deduce a formula

$$A = \Pi(B, C, \lambda J) \tag{22.5}$$

which on evaluation for $\lambda = 1$ gives a relation among the coefficients of any solution of (22.4). If B, C, J are non-negative, evaluation of the right member for $\lambda = 0$ obviously does not increase its value and also gives $\pi(B, C)$, where π is the polynomial of (22.3) and where B, C are the coefficients of U, F corresponding to the coefficients b, c of u, f. (Note in passing that we do not assume $B = 0$.) Hence we have

$$A \geqq \pi(B, C). \tag{22.6}$$

Now (22.4) is said to *dominate* (22.1) provided

(1) system (22.4) has a numerical determination with initial values $x_i = 0$,

(2) the F's are holomorphic about that numerical determination,

(3) each coefficient c of an f in (22.1) has modulus not exceeding the corresponding coefficient C in (22.4),

(4) system (22.4) has a solution which is holomorphic about $x_i = 0$ and which has all its coefficients non-negative.

From (22.3) and

$$|z_1| + |z_2| \geqq |z_1 + z_2|, \qquad |z_1|\cdot|z_2| = |z_1 z_2|$$

we find

$$\pi(|b|, |c|) \geqq |a|.$$

Since $B \geqq |b| = 0$, $C \geqq |c|$, we therefore have

$$\pi(B, C) \geqq |a|.$$

Comparison with (22.6) gives

$$A \geqq |a|.$$

Hence if (22.4) dominates (22.1), the tentative solution of (22.1) converges absolutely for every point x at which the solution of (22.4) converges absolutely.

23. Selection of a dominant system. It remains to determine a dominant system. To find a system satisfying requirements (1), (2), (3) we need only assume that S satisfies (i), (ii), (iii), (iv) of §21.

Let the functions defining f, and therefore the p, q in (22.1), be holomorphic for all x, D satisfying

$$\rho \geqq x, \qquad \rho \geqq |D|. \tag{23.1}$$

The series $p(\rho, \rho)$, $q(\rho, \rho)$ are then absolutely convergent. Let M be an upper bound for the aggregate of the terms in the series of absolute values for all $p(\rho, \rho)$, $q(\rho, \rho)$, that is, suppose

$$M \geqq | c | \rho^k, \tag{23.2}$$

where k is the total degree of the term in which c is the coefficient, for all c. Consider the fraction

$$\frac{M}{1 - \dfrac{\sum x + \sum D}{\rho}}, \tag{23.3}$$

where the set D includes all derivatives of each u_γ of order less than the corresponding h_γ.

The function given by (23.3) can be expanded by the multinomial theorem about the numerical determination $x = D = 0$. The coefficient C of any monomial of total degree k is

$$C = \frac{sM}{\rho^k}, \tag{23.4}$$

where s is a positive integer. The modulus of the corresponding term in the corresponding f is either $| c |$ or zero. In the former case, use of (23.2) in (23.4) gives

$$C \geqq | c |; \tag{23.5}$$

and a similar relation obviously holds in the latter case. Hence (23.3) furnishes a dominating function for p, q. In the case of q, however, we may subtract M since q vanishes for all its arguments equal to zero. Therefore if we take

$$F = \frac{M}{1 - \dfrac{\sum x + \sum D}{\rho}} \left(\sum E_1 + \sum E_2 + 1\right) - M, \tag{23.6}$$

we see that requirements (1), (2), (3) are satisfied.

In order to satisfy (4), we shall use both (v) and (vi) of the definition of orthonomic system. It is, moreover, necessary to modify (23.6), as will now be done.

Consider a particular equation of system (22.4). Let μ, ν_1, ν_2 be the monomials arising from the correspondents of D_1, E_1, E_2 by the substitution $x_i \rightarrow \xi_i$, $U_\alpha \rightarrow \zeta_\alpha$. If the set E_1 is not vacuous, write corresponding to each E_1 an inequality

$$\frac{\mu}{\nu_1} > 2rH_{E_1}M, \tag{23.7}$$

where H_{E_1} is defined in §22, and where r as usual denotes the number of unknowns. Note that $H_{E_1} \neq 0$ because the term E_1 is assumed present in

$\sum E_1$. If there is no term in $\sum E_1$, there is no corresponding inequality (23.7), and $\sum E_1$ can be omitted from the particular F involved.

Let inequalities (23.7) be written for every equation in (22.4). Since the first inequality system is consistent, we can find [12, §6] a solution ξ, ζ of all the inequalities (23.7) taken with $\xi_i > 1$. Now if x, E_1 are respectively multiplied by

$$\xi, \qquad \frac{\mu}{2rH_{E_1}\nu_1 M}$$

both of which exceed unity, the coefficients in (23.6) are not decreased and so continue to satisfy (23.5), although the expansion converges in a smaller region. Similarly, since the terms in E_2 are needed to dominate only terms with zero coefficients in (22.1), we may with impunity multiply the E_2 in (23.6) by any positive number (not necessarily greater than unity).

Hence, we may replace (23.6) by[4]

$$F = \frac{\mu}{1 - \dfrac{\sum \xi x + \sum D}{\rho}} \times \left(\frac{1}{2r}\sum E_1/H_{E_1}\nu_1 + \frac{1}{2r}\sum E_2/H_{E_2}\nu_2 + M\right) - M\mu. \tag{23.8}$$

24. **Reduction of the dominant system.** In (22.4) with F given by (23.8) we shall put

$$y = \sum \xi x, \qquad U_\alpha = \zeta_\alpha V_\alpha$$

and seek for the resulting system a solution V_α which involves the x's only in the combination y. We have by (12.1)

$$\frac{\partial U_\alpha}{\partial x_i} = \zeta_\alpha \frac{\partial V_\alpha}{\partial y}\frac{\partial y}{\partial x_i} = \zeta_\alpha \xi_i V'_\alpha,$$

whence

$$D_1 = \mu V_\alpha^{(h_\alpha)}, \qquad E_1 = \nu_1 V_\beta^{(h_\beta)}, \qquad E_2 = \nu_2 V_\gamma^{(h_\gamma)}.$$

All of the H_{E_1} derivatives E_1 belonging to the same unknown U_β give E_1/ν_1 the same value $V_\beta{}^{(h_\beta)}$ in the sum in (23.8). System (22.4) becomes

$$V_\alpha^{(h_\alpha)} = \frac{1}{1 - \dfrac{y + \sum D}{\rho}}\left(\frac{1}{2r}\sum_{\beta=1}^{r} V_\beta^{(h_\beta)} + M\right) - M, \tag{24.1}$$

[4] This choice of a dominating system embodies two simplifications introduced by Ritt [9, 149]: first, the coefficients are simpler in form; second, the determinant of (24.1) depends on a single parameter which Ritt chooses less than $1/r$ and which we have put equal to $1/2r$. Compare the systems in [2, 130] and [13, 297].

where $\sum D$ is a sum of positive multiples of derivatives of V_β of order less than h_β. Note that all equations belonging to the same unknown in (22.4) reduce to one equation in (24.1). Note also that (24.1) is not in solved form because all the left members appear on the right.

To put (24.1) in solved form, we write $1 - \sigma$ for the denominator (which is the same for all the equations) and sum, finding

$$\frac{1}{2r}\sum_{\beta=1}^{r} V_\beta^{(h_\beta)} = \frac{M\sigma}{1-2\sigma}. \tag{24.2}$$

Substitution in (24.1) then gives

$$V_\alpha^{(h_\alpha)} = \frac{2M\sigma}{1-2\sigma}. \tag{24.3}$$

If $(1 - 2\sigma)^{-1}$ is imagined expanded by the binomial theorem, the right member of (24.3) is seen to have non-negative coefficients.

25. **The case of ordinary equations.** To show that (24.3) has a solution with zero initial determination, we introduce sets of additional unknowns equal to $V'_\alpha, V''_\alpha, \cdots, V_\alpha^{(h_\alpha-1)}$ and thereby reduce it to a system of the first order

$$\frac{dW}{dy} = G(y, W). \tag{25.1}$$

Since (25.1) is the principal system of an orthonomic system, to prove it has a solution we need only prove that its tentative solution (1) converges and (2) satisfies the system. Here it is convenient to seek a dominant system

$$\frac{dZ_i}{dy} = \frac{M}{\left(1-\dfrac{y}{R}\right)\left(1-\dfrac{Z_1}{R}\right)\cdots\left(1-\dfrac{Z_l}{R}\right)} \tag{25.2}$$

with numerical determination different from zero.

Suppose the Z's in (25.2) all have a common value Z. Then Z must be found to satisfy

$$\left(1-\frac{Z}{R}\right)^l dZ = \frac{M\,dy}{1-\dfrac{y}{R}}$$

and to vanish for $y = 0$. Integration gives

$$Z = R - R\left[1 + \log\left(1-\frac{y}{R}\right)^{Ml+M}\right]^{1/(l+1)}.$$

The determination given the logarithm and the root can be chosen so that the right member is holomorphic [5, 356] about $y = 0$ and reduces to zero for $y = 0$. Hence we have a dominant system (25.2) for (25.1) and the tentative solution of (25.1) is known to converge. Since the coefficients of the

G's in (25.1) are non-negative and the initial determination is zero, the solution has non-negative coefficients [cf. (4) of §22].

Now if the functions

$$\frac{dW}{dy} - G(y, W) \tag{25.3}$$

are evaluated for the W's equal to their tentative expansions, holomorphic functions of y result. But the coefficients of the tentative solution were found by expressing that (25.3) and all their derivatives vanish for $y = 0$. Hence (25.3) become identically zero [5, 354] when evaluated for $y = 0$.

Accordingly, we have a solution of (25.1) which leads to a solution of (24.1) and thence to a dominant system for the system (22.1). The convergence of the tentative solution is therefore completely demonstrated for an orthonomic system.

26. **Components of an admissible system.** Separate the equations of a principal system S^* for an admissible system S into two sets S_1, $\bar{S}$, putting into S_1 all equations having x_1 for multiplier and into $\bar{S}$ all equations having x_1 for non-multiplier. Make x_1 equal to zero in the equations of $\bar{S}$. If $m = x_1^a p$, where p does not contain x_1, put

$$\left(\frac{\partial u}{\partial m}\right)_{x_1=0} = \frac{\partial v}{\partial p}, \tag{26.1}$$

where v is a new unknown given by

$$v = \left(\frac{\partial^a u}{\partial x_1^a}\right)_{x_1=0}. \tag{26.2}$$

In this way a system T, in the new unknowns v, results.

Let the principal and parametric sets for u in S be M^*, $\bar{M}$ and let the sets for the v given by (26.2) in T be N^*, $\bar{N}$. The monomial $x_1^a p$, where p is free of x_1, belongs to M^* if and only if p belongs to N'. Now the least common multiple of all monomials which are in M^* and have the form $x_1^a p$ belongs to M^* and is of the form $x_1^a q$. Hence N' contains the least common multiple of every pair of its members and $N' = N^*$.

We note that the division of $x_2, \cdots, x_n$ into multipliers and non-multipliers for $N^* + \bar{N}$ is the same as for $M^* + \bar{M}$.

If N^* is not vacuous, that is, if v is a principal unknown for T, the parametric set $\bar{N}$ consists of all divisors of monomials in N^* not contained in N^*. The set M^* is not vacuous, and $\bar{M}$ can be similarly described. If $p \in N^*$, if $q \mid p$ and if $q \,\epsilon'\, N^*$, then $x_1^a p \in M^*$, $x_1^a q \mid x_1^a p$ and $x_1^a q \,\epsilon'\, M^*$; and conversely ($p$, q meaning monomials free of x_1). Hence $\bar{N}$ is the totality of members of $\bar{M}$ divisible by x_1^a and not by x_1^{a+1}. The parametric part of v can therefore be obtained by differentiation and evaluation from the parametric part of u.

If N^* is vacuous, that is, if v is a parametric unknown[5] for T, coefficients of v may correspond to principal coefficients of u.

In the case of system $\mathfrak{S}$ a parametric unknown of the corresponding $\mathcal{T}$ arising from the second equation is

$$\left(\frac{\partial u}{\partial y}\right)_{y=0}. \tag{26.3}$$

It is clear that these terms of u involve principal coefficients and that (26.3) cannot be found from the initial determination (15.4) of $\mathfrak{S}$.

If, however, the system S^* is such that every parametric v can be determined from the initial determination for S^*, then S is called *decomposable*. Otherwise, it is *indecomposable*.

If S^* is decomposable, we repeat the process of evaluation on T. In this way, we obtain

$$S^* = S_1 + S_2 + \cdots + S_k, \tag{26.4}$$

where S_i involves the variables $x_i, x_{i+1}, \cdots, x_n$, where the unknowns in S_i are all different from those in S_j, for $i \neq j$, and where S_k is indecomposable. The systems S_i are called *components* of S. If S is indecomposable, its only component is S^*.

From the relation between $\bar{M}$ and $\bar{N}$ and the assumption about parametric v's, it is seen that the values of the functions in the initial determination for T can be found from the initial determination for S^*. Suppose, therefore, that T is determined.

Let M_1^*, $\bar{M}_1$ be the monomial sets for u and S_1. If M^* contains a monomial m not in M_1^*, that monomial must be in $\bar{M}_1$ and have the form $x_1^a p$, where p is free of x_1 and where x_1 is non-multiplier for $x_1^a p$. The function

$$\left(\frac{\partial u}{\partial m}\right)_{m^0=0}$$

which appears in the initial determination for S_1 but not in that for S^* can be found by evaluating

$$\left(\frac{\partial v}{\partial p}\right)_{m^0/x_1=0}$$

from the solution v of T. Consequently we have

THEOREM 26.1. *The initial determination for the component S_i can be found by differentiation and evaluation from the solution of the systems S_j $(j > i)$ and the initial determination for S.*

[5] I overlooked this possibility in giving a (in some respects) similar reduction for orthonomic systems [11, 72]. Theorem 49.2 of [11] is consequently true only for systems satisfying an additional condition, and assumption E is effectively proved there only for such systems.

Note that one choice of principal system for S may lead to a decomposition (26.4) with $k > 1$, whereas another may not. Note also that the order of the independent variables may affect the decomposability of a system.

The system

$$
\begin{array}{llcc}
 & & m^* & m^0 \\
 & \dfrac{\partial u}{\partial x} = v, & xz & y \\
 & \dfrac{\partial u}{\partial y} = \dfrac{\partial^2 v}{\partial z^2} + \dfrac{\partial w}{\partial y}, & yz & x \\
(26.5) & \dfrac{\partial^2 u}{\partial x \partial y} = \dfrac{\partial v}{\partial y}, & xyz & 1 \\
 & \dfrac{\partial^3 v}{\partial x \partial z^2} = \dfrac{\partial v}{\partial y} - \dfrac{\partial^3 u}{\partial z^3}, & xyz & 1 \\
 & \dfrac{\partial^2 w}{\partial x \partial y} = \dfrac{\partial^3 u}{\partial z^3} & xyz & 1
\end{array}
$$

has two components, the first component consisting of all the equations except the second and the second component consisting of the single equation

$$\frac{\partial \zeta}{\partial y} = \frac{\partial^2 \varphi}{\partial z^2} + \frac{\partial \psi}{\partial y}, \tag{26.6}$$

where

$$\zeta = u(0, y, z), \qquad \varphi = v(0, y, z), \qquad \psi = w(0, y, z).$$

The parametric unknowns in (26.6) are φ, ψ. The parametric part of v with respect to the whole system (26.5) obviously contains all terms involving y, z alone. The same statement is true of w. Hence the parametric unknowns in (26.6) can be obtained from the initial determination of (26.5) so that (26.5) is actually decomposable.

27. Definition of orderly system. A system S is *orderly* if it satisfies the six following conditions:

(i) The equations have the form

$$D' = f(x, D'').$$

(ii) The unknowns, the independent variables, the derivatives and the equations appearing explicitly are finite in number.

(iii) A numerical determination (§3) is given.

(iv) The functions f are holomorphic about the numerical determination.

(v) The first inequality system (§4) is consistent.

(vi) For some order of the independent variables and some principal system every component (§26) has a consistent second inequality system (§20).

An orderly system is admissible because it satisfies (i)–(v) of §8. Note that the components of an orderly system are orthonomic. If (vi) of §21 is

satisfied by S, it will also be satisfied by the components of S. This is obvious for S_1 because it is a subset of S. Consider an inequality of (20.5) for T (§26)

$$\omega_i + \cdots \geqq \omega_{ji} + \cdots . \tag{27.1}$$

Any index i in it occurs just once on the left and just once on the right in the second position. If the i-th equation of T belongs to the v in (26.1), there is for S the corresponding condition

$$(\omega_i + a) + \cdots \geqq (\omega_{ji} + a) + \cdots . \tag{27.2}$$

As far as the index i is concerned, (27.2) is equivalent to (27.1). Since the statement applies to all indices, (27.1) and (27.2) are equivalent. Thus the second inequality system for S implies that for T. Hence every orthonomic system is orderly.

Next we have to prove that the tentative solution of §18 converges for every orderly system. We employ the decomposition (26.4). Since S_k is orthonomic, it has a convergent tentative solution. A tentative solution of S_{k-1} can be determined so as to reduce to the initial determination given by the initial determination for S and the convergent tentative solution of S_k (Theorem 26.1). Proceeding in this way, we construct a convergent tentative solution of S_1, which must coincide with the tentative solution of S found in §18. Hence the tentative solution of S converges and we have

THEOREM 27.1. *The expansions of the tentative solution for an orderly system always converge and therefore define functions holomorphic about the initial values.*

28. **Passivity conditions.** Corresponding to each equation (2.4) of S^* for an orderly system write the equation

$$D' - f = F, \tag{28.1}$$

and give F the multipliers and non-multipliers of D'. We shall refer to F as the "equation F" of S. From §10 we know that when the tentative solution or any other set of holomorphic u's is put in (28.1), F becomes a holomorphic function of the x's. When F is once so determined, corresponding to (28.1) there is a relation

$$D' = f + F, \tag{28.2}$$

which is identically satisfied in the x's when the u's are replaced by the holomorphic functions defining F.

The conditions used to determine the coefficients of the tentative solution were essentially

$$\left(\frac{\partial F}{\partial m}\right)_{m^0=0} = 0, \tag{28.3}$$

where F belongs to S^*, where m is a monomial in the multipliers of F and where m^0 is the set of non-multipliers of F. Since m, m^0 have no x in common, (28.3) becomes

$$\frac{\partial}{\partial m}[(F)_{m^0=0}] = 0. \tag{28.4}$$

Hence F computed for u's constituting a tentative solution vanishes identically in the multipliers when the non-multipliers are made zero. We wish next to find the conditions that F vanish for all values of the non-multipliers.

Let x_0 be a non-multiplier for (28.1). If $D' \sim m'$, then $m'x_0$ is in the principal set M^*. Hence there is an equation in S^* with left derivative $E' \sim m'x_0$. Write it in the form (28.1) as

$$E' - g = G. \tag{28.5}$$

Differentiation of (28.2) and substitution from the result and (28.5) in

$$\frac{\partial D'}{\partial x_0} = E'$$

give [cf. (12.1) for the meaning of δ_0]

$$\frac{\partial F}{\partial x_0} = G + g - \delta_0 f. \tag{28.6}$$

(This equation can be interpreted in two useful ways: (i) as a relation between the functions defined by (28.1), (28.5) or (ii) as saying that the result of differentiating the equation $D' = f$ and subtracting the equation $E' = g$ is the equation $\delta_0 f = g$.) Moreover, if F in S^* and G in S are distinct equations (§2) having the same left derivative we get a relation of the form

$$F = G + g - f. \tag{28.7}$$

Just as in the case of the extended systems (§18), the principal derivatives can be eliminated from $g - \delta_0 f$ and $g - f$ by means of the identities (28.2) and the results of differentiating them. In this way, (28.6) and (28.7) are converted into

$$\frac{\partial F}{\partial x_0} = H(x, F) + K(x, D''), \tag{28.8}$$

$$F = H(x, F) + K(x, D''), \tag{28.9}$$

where H is a homogeneous polynomial in the functions F and their derivatives and K involves only *parametric derivatives* D''.

Since from homogeneity

$$H(x, 0) = 0,$$

a necessary condition that the tentative solution satisfy S is

$$K(x, D'') = 0. \tag{28.10}$$

As we are seeking a solution in which the parametric derivatives have arbitrary initial values, (28.10) must be identically satisfied in all the arguments.

The totality of conditions (28.10) formed for all choices of an equation F and a pertinent non-multiplier x_0 and for all choices of an equation in S but not in S^* is called the *passivity conditions* of S. If the passivity conditions are identically satisfied in all their arguments, S is *passive*.

It is convenient to repeat here the exact process for forming the passivity conditions.

Let F be an equation of the principal system S^* (p. 261) *and let x_0 be a non-multiplier* (p. 256) *for the left derivative of F. The left derivative of $\dfrac{\partial F}{\partial x_0}$ is the left derivative of another equation G in S^*. Eliminate principal derivatives* (p. 255) *from the difference $\dfrac{\partial F}{\partial x_0} - G$. The resulting equation in independent variables and parametric derivatives is one of the passivity conditions for every choice of F and x_0.*

Let F be an equation in the principal system S^* and G a distinct equation in S with the same left derivative. Eliminate from $F - G$ the principal derivatives. The resulting equation in the independent variables and parametric derivatives is one of the passivity conditions for every choice of F and G.

Denote the equations of $\mathfrak{S}^*$ by F_i ($i = 1, \cdots, 5$), and the equation omitted from $\mathfrak{S}$ in forming $\mathfrak{S}^*$ by F_6. Differentiating each equation with respect to each of its non-multipliers and eliminating principal derivatives give

$$\frac{\partial F_1}{\partial x} = -F_1 + F_2 - F_3,$$

$$\frac{\partial F_1}{\partial y} = F_3,$$

$$\frac{\partial F_2}{\partial y} = F_4,$$

$$\frac{\partial F_3}{\partial x} = -F_3 + F_4 - \frac{\partial F_3}{\partial y},$$

$$F_6 = -F_2 + F_3 + F_5.$$

Since all the parametric derivatives cancel, the system is passive.

29. **Passive systems.** Suppose S passive. The relations (28.8), (28.9) then become

$$\frac{\partial F}{\partial x_0} = H(x, F), \tag{29.1}$$

$$F = H(x, F). \tag{29.2}$$

The totality of these relations can be regarded as a system T for the determination of the F's corresponding to the tentative solution of S.

To prove that T is an admissible system it is only necessary to show that its first inequality system is consistent. From (28.6) we see that (29.1) contributes inequalities of the types

$$Fx_0 > G, \qquad Fx_0 > H, \tag{29.3}$$

where H in (28.1) represented an equation used in eliminating principal derivatives from $g - \delta_0 f$ and where F, G, H, x_0 now represent, of course, the unknowns in the inequality system. Suppose the equations of S^* corresponding to F, H have left derivatives corresponding respectively to m', p'. Then G has left derivative corresponding to $m'x_0$. Introduce into (29.3) new unknowns by means of the definitions

$$F = m'F', \qquad G = m'x_0G', \qquad K = p'K'.$$

The inequalities (29.3) become

$$F' > G', \tag{29.4}$$

$$m'x_0 > \frac{K'}{F'}p'. \tag{29.5}$$

System (29.4) is consistent. To see this, let the order of the left members of equations F, G be F'', G''. Then

$$G'' > F''. \tag{29.6}$$

Consequently $-F''$, $-G''$ is a solution of (29.4). Addition of a large positive integer will give a positive solution.

Exactly the same types of inequalities are obtained from the consideration of (29.2). The inequalities like (29.4) have the form

$$F' > L', \tag{29.7}$$

where F belongs to S^* and L belongs to S but not to S^*. Relations like (29.6) are not available because F and L have equal left derivatives. Supposing a solution of (29.4) at hand, however, we can easily satisfy (29.7) because all the unknowns are on the right. Put with (29.5) the inequalities of the second type yielded by (29.2). Since the inequality

$$m'x_0 > p'$$

belongs to the first inequality system for S, system (29.5) is consistent [12, §6]. From the solution of (29.4), (29.5), (29.7) we immediately deduce a solution of the inequality system for system T, which is accordingly seen to be admissible.

The parametric terms in the expansion of the unknown F of system (29.1) are the terms in the multipliers which F has with respect to (28.1). Equation (28.4) therefore says that the initial determination of (29.1) is zero. The system (29.2) has vacuous initial determination because all the coefficients are principal. Hence (29.1), (29.2) together with their zero initial determination

constitute a determined admissible system. This system is obviously satisfied by making all the F's equal to zero. Theorem 18.1 shows that the F's must be zero, and we have

THEOREM 29.1. *A passive orderly system has a unique holomorphic solution corresponding to each arbitrarily given relevant initial determination.*

This theorem can be used to prove that an orderly system which is passive for one choice of the principal system S^* is passive for all.

30. **Sequences of monomials.** Before discussing non-passive systems, we need to develop several preliminary results. The first of these is [8, 554], [2, 68]:

THEOREM 30.1. *Every sequence of monomials in which no monomial is a multiple of any of those preceding it is necessarily finite.*

The theorem is obvious in the case of a single variable. The sequence contains at most $a + 1$ monomials x^a, x^{a-1}, $\cdots$, x, 1.

In the case of n variables, let $x_1^{i_1}x_2^{i_2} \cdots x_n^{i_n}$ be the first monomial of the sequence. Denote by E_{jk} the set of all monomials of the sequence in which x_j has the exponent k. Since for every monomial of the sequence the exponent of at least one x does not exceed the exponent of that x in the first monomial, if k ranges from 0 to i_j, for each j, every monomial of the sequence is contained in at least one set E_{jk}. The monomials in each E_{jk} can be thought of as arranged in the same relative order as in the original sequence. Except for a common factor x_j^k, each E_{jk} is a sequence of monomials in $n - 1$ variables such that no one of them is a multiple of any of the preceding. If the theorem is assumed for $n - 1$ variables, it follows that there is but a finite number of monomials in each E_{jk}. The number of sets E_{jk} being finite, the theorem also holds for n variables, and the induction is complete.

31. **Complete ordering of sets of monomials.** The monomials in the x's will be called *completely ordered* if values greater than unity have been assigned to the x's so that the equation

$$x_1^{i_1}x_2^{i_2} \cdots x_n^{i_n} = x_1^{j_1}x_2^{j_2} \cdots x_n^{j_n}$$

for integral i, j implies

$$i_1 = j_1, \ i_2 = j_2, \ \cdots, \ i_n = j_n .$$

Finding x's completely ordering the monomials is equivalent to finding values for $\log x_1$, $\log x_2$, $\cdots$, $\log x_n$ so that

$$k_1 \log x_1 + k_2 \log x_2 + \cdots + k_n \log x_n = 0$$

for integral k's implies

$$k_1 = k_2 = \cdots = k_n = 0.$$

One way of completely ordering the monomials is to set $\log x_i = \pi_i$, where π_i is the positive square root of the i-th positive prime. To see that such is the case, we need to know the linear independence of the π's with respect to the ring of integers. This is a consequence of

THEOREM 31.1. *If $\sigma_1, \sigma_2, \cdots, \sigma_k$ are k distinct numbers from the set $\pi_1, \pi_2, \cdots, \pi_n$, no σ_i is contained in the field obtained by adjoining the $k - 1$ others to the rational field.*

The theorem is trivial for $k = 1$. Assume it is true for $1, 2, \cdots, k - 1$. Denote by an R with subscripts the field obtained by adjoining to the rational field all σ's except those having the subscripts on the R. Suppose σ_k is in the field R_k. Since the squares of all σ's are rational, we have

$$\sigma_k = a\sigma_{k-1} + b \qquad (a \neq 0), \tag{31.1}$$

where a, b are in $R_{k,k-1}$. Squaring, we find that

$$a^2\sigma_{k-1}^2 + 2ab\sigma_{k-1} + b^2$$

equals a rational number. If $b \neq 0$, then σ_{k-1} is in $R_{k,k-1}$. Hence $b = 0$ and a^2 is rational. Adjusting the notation, if necessary, we may put

$$a = c\sigma_{k-2} + d \qquad (c \neq 0),$$

where c, d belong to $R_{k,k-1,k-2}$. Reasoning as before gives $d = 0$ and c^2 rational. Continuing we finally have

$$\sigma_k = e\sigma_{k-1}\sigma_{k-2} \cdots \sigma_1,$$

where e is rational. Suppose $e = p/q$ with $(p, q) = 1$. Then

$$q^2\sigma_k^2 = p^2\sigma_{k-1}^2\sigma_{k-2}^2 \cdots \sigma_1^2. \tag{31.2}$$

The prime σ_k^2 divides p^2 and therefore p. Hence its square divides the right side of (31.2) and therefore q. Since this contradicts the assumption $(p, q) = 1$, the theorem is proved.

As in §5, the substitution of these values for x in monomials m_1, m_2 gives ordinals k_1, k_2. For a complete ordering, however, we can say that m_1 follows, coincides with or precedes m_2 according as $k_1 > k_2$, $k_1 = k_2$, $k_1 < k_2$.

As an example, let there be three independent variables x, y, z and put in accordance with the above

$$\log x = \sqrt{2}, \qquad \log y = \sqrt{3}, \qquad \log z = \sqrt{5}. \tag{31.3}$$

The monomials of the second degree are ordered as follows:

$$z^2 > yz > xz > y^2 > xy > x^2. \tag{31.4}$$

To verify this, write the corresponding exponents:

$$2\sqrt{5} > \sqrt{3} + \sqrt{5} > \sqrt{2} + \sqrt{5} > 2\sqrt{3} > \sqrt{2} + \sqrt{3} > 2\sqrt{2}.$$

If we vary the choice by setting

$$\text{(31.5)} \qquad \log x = \sqrt{2}, \qquad \log y = \sqrt{5}, \qquad \log z = \sqrt{7},$$

the ordering is slightly changed:

$$\text{(31.6)} \qquad z^2 > yz > y^2 > xz > xy > x^2.$$

The two orderings (31.4) and (31.6) are the only ones possible, if the linear monomials are in the relation $z > y > x$.

Although stated above for pure monomials, this method of ordering applies to the monomials $u_\alpha x_1^{i_1} x_2^{i_2} \cdots x_n^{i_n}$: put, for example, $\log u_\alpha = \pi_{n+\alpha}$. In such an ordering we have

$$\log u_\alpha - \log u_\beta + k_1 \log x_1 + k_2 \log x_2 + \cdots + k_n \log x_n = 0$$

for derivatives of equal ordinal. We conclude $1 = -1 = 0$ unless $\alpha = \beta$. Hence two derivatives with equal ordinals are identical.

In §33, however, we do not need such an à priori ordering of the unknowns. Rather, we let their ordering be determined from the inequality system in the manner described in the next section.

The ordering described in this section arranges all derivatives of each unknown in a definite order, which is the same for all unknowns. This ordering of the derivatives of a single unknown does not have the property:

$$\text{(31.7)} \qquad \text{if } D_1 \text{ is of higher order than } D_2\text{, then } D_1 > D_2.$$

For example, (31.5) imply $z^2 > x^3$. In this respect, the ordering differs radically from an ordering by cotes even when the cotes are allowed to be irrational. As will be done in §33, however, from any complete ordering O_1 can be deduced another complete ordering O_2 which makes (31.7) valid for any given *finite* set of pure monomials and which establishes the same relative order as O_1 between every pair of monomials of equal degree. Thus, if the base is chosen as 2, multiplying x, y, z given by (31.5) by 4 preserves the ordering (31.6) but makes x^3 follow z^2 because $2 > 2\sqrt{7} - 3\sqrt{2}$. But for no ordering by the present scheme does (31.7) hold for *all* monomials: if $z > x > 1$, there are always infinitely many positive integers a satisfying $z^a > x^{a+1}$.

32. **Completion of ordering.** A consistent first inequality system at least partly orders the derivatives. We now wish to show how values satisfying the first inequality system and at the same time completely ordering the derivatives can be found.

Suppose $x_{k+1}, \cdots, x_n$ have been determined so as to satisfy the inequalities which are implied by the first inequality system and which involve only $x_{k+1}, \cdots, x_n$ [12]. Suppose further that the partial solution has the form

$$\text{(32.1)} \qquad (\log x_i)^2 = \frac{p_i}{q_i} \qquad (i = k+1, \cdots, n),$$

where the p's are prime, and where every pair from $p_{k+1}, \cdots, p_n, q_{k+1}, \cdots, q_n$ is relatively prime. Then no fraction p_i/q_i is the square of a rational number.

In the solution of the first inequality system x_k is found to lie on a certain segment [12]. Let s be a positive irrational number of the segment on which $(\log x_k)^2$ is constrained to lie. Let a positive ϵ be chosen so that every x_k satisfying

$$s + \epsilon > (\log x_k)^2 > s - \epsilon \tag{32.2}$$

is an admissible value for x_k. Determine for s an approximation p/q, where p, q are positive integers satisfying

$$\tfrac{1}{2}\epsilon > \left|\frac{p}{q} - s\right|, \qquad \tfrac{1}{2}\epsilon q > |1 - s|. \tag{32.3}$$

Use of the identity

$$\frac{Nq}{Nq+1}\left(\frac{p}{q} - s\right) + \frac{1-s}{Nq+1} = \frac{Np+1}{Nq+1} - s$$

shows that inequalities (32.3) imply

$$\epsilon > \left|\frac{Np+1}{Nq+1} - s\right| \tag{32.4}$$

for every positive integer N.

The arithmetic progression whose $(r + 1)$-th term is

$$rp_{k+1} \cdots p_n q_{k+1} \cdots q_n p + 1 \tag{32.5}$$

contains an infinite number of primes. Let r be fixed so that (32.5) is a prime exceeding $|p - q|$. Set

$$N = rp_{k+1} \cdots p_n q_{k+1} \cdots q_n.$$

The positive integers p_k, q_k defined by

$$p_k = Np + 1, \qquad q_k = Nq + 1 \tag{32.6}$$

are relatively prime to N. A prime dividing p_k and q_k would have to divide $p_k - q_k = N(p - q)$. Hence $(p_k, q_k) = 1$. Moreover, p_k/q_k is not the square of a rational number since its numerator is prime.

From (32.4) we see that we may choose

$$(\log x_k)^2 = \frac{p_k}{q_k},$$

in constructing a solution of the first inequality system. In this way, (32.1) is extended to include one more value of the index i. We may accordingly assume that (32.1) for $k = 0$ gives a solution of the first inequality system. Moreover, Theorem 31.1 is readily seen to remain true if σ_i-log x_i is defined by

(32.1) rather than by π_i as in §31. Equation (31.2), for example, becomes on the substitution $\sigma_i^2 = p_i/q_i$

$$q^2 p_k q_1 \cdots q_{k-1} = p^2 p_1 \cdots p_{k-1} q_k ,$$

and the prime p_k must divide both p and q.

33. **Reduction to passive form.** Consider now a system T which is not in the solved form (2.4). Let the monomials in the x's be completely ordered. Suppose no derivative (§2) of $u_{k+1}, \cdots, u_r$ occurs in T, but that a derivative of u_k does. Consider the set of all derivatives of u_k which appear as arguments in T. Among them consider those of maximum order. The derivative whose monomial follows all the others in this subset is called the *follower* of T. We shall consider systems satisfying the following conditions:

(A) The equations have the form

$$f(x, D) = 0.$$

(B) The unknowns, the independent variables, the derivatives appearing and the equations are finite in number.

(C) A numerical determination (§3) is given.

(D) The functions f are holomorphic about the numerical determination.

(E) The monomials in the independent variables have been completely ordered (§31).

(F) The system has a follower (§33).

(G) The partial derivative of at least one f with respect to the follower does not vanish for the numerical determination.

Such a system will be called *manageable*. It is always equivalent to a system T_1 which contains just one equation and which satisfies

$$\omega_i \geqq \omega_{ji} , \tag{33.1}$$

where i refers to T_1 and j to T, plus a system T_2, whose follower if existent is different from that of T. To see this, select one of the equations referred to in (G), solve it by the implicit function theorem [6, 12] for the follower to get T_1, and substitute the result in the other equations to get T_2. System T_2 has properties (A) to (E). If it also has properties (F), (G), the operation can be repeated.

If only manageable systems are encountered, repetition of the elimination gives a system S which satisfies (i)–(iv) of §21, and also (vi) because it satisfies (33.1). We can show that its first inequality system is consistent in the following way. For convenience, let the followers in the above reduction belong to $u_r, u_{r-1}, \cdots, u_{l+1}$. All the inequalities contributed to the first system by equations belonging to u_r have u_r on the left. Those which also have u_r on the right are of two types, which we shall designate by H, L. Those of type H have both left and right member of the same degree. They are automatically

satisfied by the ordering of the x's. Those of type L have left member of higher degree than the right. They can all be satisfied [12, §5] by multiplying all the x's by a factor. The others can all be satisfied by taking u_r large enough, once the inequalities contributed by equations belonging to $u_{r-1}, \cdots, u_{l+1}$ have been satisfied. As the last mentioned do not involve u_r, the first inequality system for S has a solution. Hence S is orthonomic (§21).

Let the passivity conditions of S be formed. The resulting system $\bar{T}$ has properties (A), (B), (E). It has for arguments parametric derivatives of S and the independent variables. Suppose that the numerical determination of S can be extended so as to give a numerical determination for $\bar{T}$ and that it is possible to apply the solution process outlined above for T to $\bar{T}$ without an unmanageable system being encountered.

In this way, it may be possible to construct a sequence

$$S_1, \quad S_2, \quad S_3, \quad \cdots \tag{33.2}$$

in which the S's are all orthonomic and S_{i+1} expresses the passivity conditions of S_i. Let the monomial sets for S_i be $M_{\alpha i}$ and consider a fixed α. Since the left derivatives in S_{i+1} are parametric in S_i, no monomial in $M_{\alpha,i+1}$ is divisible by a monomial in $M_{\alpha,i}$ and the sequence (33.2) must be finite (Theorem 30.1).

Accordingly, there are three possibilities.

(1) An unmanageable system may be encountered or a numerical determination may not be capable of extension, so that sequence (33.2) is not defined.

(2) Sequence (33.2) is defined and its last member contains an equation which involves at most the independent variables and which is not identically satisfied.

(3) Sequence (33.2) is defined and its last member contains only equations identically satisfied.

In case (2), the proposed system has no solution holomorphic about the numerical determination.

In case (3), it is easy to see that (33.2) constitutes a special passive orthonomic system, which has a solution of the type specified in Theorem 29.1.

Except for the satisfaction of (33.1) and the absence of the cotes, the reduction just given is Riquier's [8, 556]. Because of possibility (1) it has limitations, as Riquier himself was aware [8, 556–557]. Nevertheless, the reduction not only is very useful in particular cases but also seems destined to remain the guiding principle in the formulation of processes whose domain of applicability is more clearly indicated in advance. The first to seek such a process was Ritt who [9] has given a highly satisfactory elimination theory, which insures the existence of a sequence (33.2) when the functions f are polynomials in the D's with coefficients holomorphic in the x's. Subsequently the writer [11, 75–77] gave for non-algebraic systems a reduction process conditioned by our ability to find numerical determinations.

34. **The complete and complementary sets.** The principal and parametric sets of §13, introduced in [11, 56] under different names, seem best adapted to

the proof of the existence theorem. In practice, however, they are conveniently replaced by the older complete and complementary sets into which they can be collapsed. The initial determination and the passivity conditions can be more compactly, although less symmetrically, expressed in terms of the contracted sets.

We suppose the independent variables x_i ordered, say by increasing subscript. If two distinct monomials m, p are given by

$$m = x_1^{i_1}x_2^{i_2} \cdots x_n^{i_n}, \qquad p = x_1^{j_1}x_2^{j_2} \cdots x_n^{j_n},$$

then m is said to be *higher* or *lower* than p according as the first non-zero difference in the sequence

$$i_1 - j_1, i_2 - j_2, \cdots, i_n - j_n,$$

read from right to left, is positive or negative. In this way, the monomials are ordered according to *rank*.

Consider in the set $M = M^* + \bar{M}$ of §13 a pair of monomials p, m satisfying the two conditions

(1) $p = mx_0$, where x_0 is non-multiplier for m;

(2) the set of multipliers for p is that for m augmented by x_0.

The two terms of the finite sum (14.2) corresponding to m, p, say

$$mb + pc,$$

can be written as one

$$md,$$

where d is a function of the multipliers of p.

Let m be the monomial of minimum rank in M admitting a partner p which satisfies (1), (2). If p does not belong to the original set M' from which M was constructed, omit p from M and compensate by giving m the additional multiplier x_0. Repeat the operation as long as the diminished set has a pair m, p with p not in M'. In this way, for the given order of the independent variables, there is obtained a unique set $M_R = M_1 + M_2$, where M^*, $\bar{M}$ have collapsed into M_1, M_2 respectively. M_1 contains the distinct monomials in M' and possibly some multiples of them. The sets M_1, M_2 were called by Janet [2, 68–91] respectively the *complete* and *complementary sets*. Riquier [8, 130] had previously employed the equivalent of M_2 in the description of the parametric part of the development, which he called the residue. Riquier seems nowhere to make effective use of the representation of the principal part as a finite sum, although except for interpretation he is in possession of the formal results. Compare, for example, the process [8, 166], which he only stated for the parametric part, with the processes of §§14, 19. A unique process for obtaining each principal coefficient by means of the complete set without con-

sidering the question of passivity seems Janet's contribution. This process in particular makes it possible for him to formulate for orthonomic systems, as is done here in §28 for orderly systems, passivity conditions which are much simpler than Riquier's [8, 357], although equivalent to them.

For M_1, M_2, Theorems 13.1, 13.2 need to be modified slightly so as to read [11, 58]

THEOREM 34.1. *The product of a monomial p by one of its non-multipliers is equal to the product of a unique monomial m of the complete set by multipliers alone. The monomial m is of higher rank than p.*

The proof is immediate.

If in the discussion of the existence theorem the sets M^*, $\bar{M}$ are interpreted respectively as the complete and complementary sets rather than the principal and parametric, everything developed so far holds as given, except for one detail: when equations (28.6) and the passivity conditions (28.10) are formed, derivatives of the functions G, g appear rather than the functions themselves. Instead of finding the left derivative of $\partial F/\partial x_0 \sim um$ as the left derivative of some G, we find the $G \sim up$, where p is the generator of m, and differentiate G with respect to multipliers so as to produce $\partial F/\partial x_0$. This, however, does not affect the proof. The passivity conditions are effectively the same except for condensation corresponding to that in the sets M.

The collapse of the principal and parametric sets for u into the complete and complementary sets is illustrated below for system $\mathfrak{S}$.

00–00	00–01	00–01
*10–00	*10–00	*10–10
20–10	20–10	
01–01		
11–01	11–01	11–01
*21–11	*21–11	*21–11

The condensation in the Maclaurin expansion is as follows (cf. §14).

$$u = 1b_{00}(y) + xb_{10}(x) + xyb_{11}(y) + x^2yb_{21}(x, y).$$

In the notation of §28, the corresponding system $\mathfrak{S}^*$ is

		m^*	m^0
F_1	$\dfrac{\partial u}{\partial x} = \dfrac{\partial u}{\partial y} + u,$	x	y
F_3	$\dfrac{\partial^2 u}{\partial x\,\partial y} = \dfrac{\partial^2 u}{\partial y^2} + \dfrac{\partial u}{\partial y},$	y	x
F_4	$\dfrac{\partial^3 u}{\partial x^2\,\partial y} = \dfrac{\partial^3 u}{\partial y^3} + 2\dfrac{\partial^2 u}{\partial y^2} + \dfrac{\partial u}{\partial y},$	xy	1
F_6	$\dfrac{\partial v}{\partial y} = \dfrac{\partial^3 u}{\partial y^3} + \dfrac{\partial u}{\partial y}.$	xy	1

Expressing the passivity conditions, we find the relations (cf. §28)

$$\frac{\partial F_1}{\partial y} = F_3,$$

$$\frac{\partial F_3}{\partial x} = -F_2 + F_4 - \frac{\partial F_3}{\partial y},$$

$$F_6 = -\frac{\partial F_1}{\partial x} - F_1 + F_5.$$

A more complicated example follows. Let M' be

$$xy, \quad xz, \quad z^2.$$

000-000	000-100	000-100	000-100	000-100	000-100	000-100
100-100						
010-010	010-010	010-010	010-010	010-010	010-010	010-010
*110-110	110-110	110-110	110-110	110-110	110-110	110-110
001-000	001-000	001-010	001-010	001-010	001-010	001-010
*101-100	101-100	101-100	101-110	101-110	101-110	101-110
011-010	011-010					
111-110	111-110	111-110				
*002-001	002-001	002-001	002-001	002-101	002-101	002-111
102-101	102-101	102-101	102-101			
012-011	012-011	012-011	012-011	012-011	012-111	
112-111	112-111	112-111	112-111	112-111		

Consider the equation

$$F = \frac{\partial^2 u}{\partial x\, \partial y} - f$$

whose non-multiplier is z. The derivative $\partial^3 u/\partial x \partial y \partial z$ is not a left derivative. But there is an equation

$$G = \frac{\partial^2 u}{\partial x \partial z} - g,$$

and xz is the generator of xyz. Hence the passivity condition corresponding to F and z is formed by differentiating G with respect to its multiplier y:

$$F_z - G_y = \delta_y g - \delta_z f.$$

In practice, it is not necessary to rewrite the unchanged monomials as was done above for the sake of clarity. The reduction can be performed by simply striking out monomials and modifying multipliers in the first column. Rules for obtaining the complete and complementary sets directly from M' without using M are given by Janet [2, 80, 89]. The above method is regarded as simpler because it determines both sets simultaneously by a single process.

A method of finding the sets directly from the Maclaurin series is given in [13, 286], previously for the complementary set in [8, 131–135]. It is easy to apply in the simpler cases. The idea behind it is to manipulate the Maclaurin series into a finite sum (14.2) so that each monomial of the given set appears multiplied by a b.

35. The old and new definitions of orthonomic system. The present section establishes the exact connection between Riquier's definition[6] of orthonomic system and the one given in §21. The crux of the matter is to prove the following theorem and its converse.

THEOREM 35.1. *Let S satisfy* (i)–(iv) *of* §21. *If S satisfies in addition* (v) *and* (vi), *there exist cotes ordering the derivatives of S so that each left derivative follows every corresponding right derivative.*

We shall try to determine for the unknowns u_α first cotes c_α so that the first cote of each left derivative is at least equal to the first cote of every corresponding right derivative.

In §7 we agreed to regard a right member as containing a derivative of u_α of order $-\infty$ if no derivative of u_α is actually present. Such a derivative is to be given a first cote $c_\alpha - \infty$. Let α_i be the index of the unknown to which the i-th equation belongs. The conditions on the c's are then

$$\omega_i + c_{\alpha_i} \geqq \omega_{ij} + c_{\alpha_j} \qquad (i, j = 1, 2, \cdots, \kappa). \tag{35.1}$$

Superficially, these inequalities appear to put greater restrictions on the unknowns than (20.4) because the unknowns are in general fewer in (35.1) than in (20.4). This difference, however, is compensated by the relations (7.1).

The elimination of the unknowns as in [12, §7] leads to the conditions of consistency

$$\sum (\omega_i - \omega_{ij}) \geqq 0, \tag{35.2}$$

where i runs through an arbitrary cycle and j has the value immediately following that of i in the cycle. Reversing the sense of the cycle is equivalent to interchanging the indices on the ω. Hence (35.2) are equivalent to (20.5). First cotes can be assigned in the required manner if and only if the second inequality system is consistent.

Suppose first cotes have been assigned so as to satisfy (35.1), and consider the totality of order relations

$$D_1 > D_2, \tag{35.3}$$

where D_1 is a left derivative of S and D_2 a corresponding right derivative. These relations fall into two categories: first, those which are actually established by the first cotes, and second, those for which D_1 and D_2 have equal first cotes. Only the latter need have our attention at the moment. They are not established by the first cotes.

Fix upon a solution of the first inequality system such that $\log u_\alpha$, $\log x_i$ are integers [12, §6].[7] Take these integers as second cotes. If the D_1, D_2 in (35.3) correspond to m_1, m_2 respectively, then $\log m_1$, $\log m_2$ are the second cotes of

[6] For the comparisons of this section it is convenient to refer to [13, 303].

[7] The restriction that the cotes be integers is not essential. If the first cotes of the independent variables are made unity and all the other cotes are allowed to assume any values in the real field, an equivalent theory is obtained.

D_1, D_2. Since $m_1 > m_2$, we have $\log m_1 > \log m_2$ and the second cotes indicated establish the desired order relations.

Conversely, if there are cotes establishing the order relations (35.3), then $m_1 > m_2$ for each relation (35.3). To see this, it is easiest to take $\epsilon = 1$ in the application of Theorem 3 of [13, 291]. The first inequality system is therefore consistent.

A comparison of our definition with Riquier's accordingly shows that the two are essentially equivalent. To make them identical it is necessary to add the following two requirements to those already given in §21.

(vii) The left derivatives of distinct equations are distinct.

(viii) No left derivative is a right derivative.

These are restrictions of convenience which do not limit the generality of the results. We do not make them because in our treatment it is advantageous to omit them.

We note in passing that (ii) and (iv) of [13, 303] imply that the number of derivatives is finite, essentially because the order is bounded.

36. **A non-orthonomic orderly system.** The system (26.5) has first inequality system

$$ux > v, \quad uy > vz^2, \quad u > w, \quad xz^2 > y, \quad vx > uz, \quad wxy > uz^3,$$
$$x > 1, \quad y > 1, \quad z > 1,$$

which is consistent because it has the solution

$$u = 5, \qquad v = w = 1, \qquad x = 24, \qquad y = z = 2.$$

The order matrix (§7) is

$$\begin{array}{c|ccccc} 1 & -\infty & -\infty & -\infty & 0 & -\infty \\ 1 & -\infty & -\infty & -\infty & 2 & 1 \\ 2 & -\infty & -\infty & -\infty & 1 & -\infty \\ 3 & 3 & 3 & 3 & 1 & -\infty \\ 2 & 3 & 3 & 3 & -\infty & -\infty \end{array}.$$

The relation

$$\omega_2 + \omega_4 \geqq \omega_{24} + \omega_{42}$$

of the consistency conditions (20.5) for the second inequality system is not satisfied. Hence the system is not orthonomic. On the other hand, if the second equation is omitted, it is readily verified that all the conditions are satisfied by the remaining system. In §26 we saw that (26.5) is decomposable and that its first component S_1 is obtained by omitting the second equation. Hence S_1 is orthonomic. It is readily seen that S_2 is also orthonomic. Hence S is orderly but non-orthonomic.

Since the second inequality system for S is inconsistent, first cotes cannot

be found for the unknowns, as is also easily seen directly. Riquier's theorem [13, 301] for non-orthonomic systems does not apply.

37. **Two non-orderly systems.** To show that the tentative solution does not always converge Kowalevsky [4, 22] employed an example which has become classical. In our terminology, she showed that the determined admissible system (15.6) does not have a holomorphic solution. The system satisfies (i)–(v) of the requirements for either orderly or orthonomic systems, but (vi) for neither. Accordingly, a tentative solution exists; but it does not converge. Because (vi) is not satisfied it is impossible to use the form of dominating system employed in §22. It is, however, interesting to see why another form of system, say

$$\frac{\partial U}{\partial x} = \frac{3}{1 - 2\left(x + y + \dfrac{\partial^2 U}{\partial y^2}\right)} - 3, \tag{37.1}$$

fails to work. The right member of (37.1) dominates that of (15.7) because we may take $M = 3$, $\rho = \frac{1}{2}$. Seeking a solution which is a function of $z = x + y$ leads to the equation

$$U' = \frac{3}{1 - 2(z + U'')} - 3. \tag{37.2}$$

This equation has a holomorphic solution with $U(0) = U'(0) = U''(0) = 0$. It is found, however, that

$$U'''(0) = -1.$$

The coefficient of y^3 in the initial determination of the solution of (37.1) furnished by (37.2) is accordingly negative. Consequently this U does not satisfy requirement (4) of §22.

It is easy to show that the tentative solution converges only for $x = y = 0$.

A somewhat different situation arises in connection with another classical example

$$p_{11} = f(p_{20}, p_{02}), \tag{37.3}$$

where

$$p_{ij} = \frac{\partial^{i+j} u}{\partial x^i \partial y^j}.$$

Differentiation gives

$$p_{12} = f_1 p_{21} + f_2 p_{03}, \qquad p_{21} = f_1 p_{30} + f_2 p_{12}.$$

If

$$f_1 f_2 = 1,$$

for example if

$$f = p_{20} + p_{02},$$

we find

$$f^1_1 p_{30} + f^1_2 p_{03} = 0,$$

that is, a relation among the parametric derivatives. Accordingly, (37.3) does not have a solution corresponding to arbitrary initial determination. Since (37.3) satisfies (vi) but not (v) of §27, we see that (v) plays an essential rôle not only in the convergence proof but also in the determination of the coefficients in the discussion as we have developed it.

References

1. M. Janet, *Leçons sur les Systèmes d'Equations aux Dérivées Partielles*, Paris, 1929.
2. M. Janet, *Les systèmes d'équations aux dérivées partielles*, Journal de Mathématiques, (8), vol. 3(1920), pp. 65–151.
3. M. Janet, *Les Systèmes d'Equations aux Dérivées Partielles*, Mémorial des Sciences Mathématiques, fasc. 21, Paris, 1927.
4. S. Kowalevsky, *Zur Theorie der partiellen Differentialgleichungen*, Journal für Mathematik, vol. 80(1875), pp. 1–32.
5. W. F. Osgood, *Lehrbuch der Funktionentheorie*, vol. I, 5th edition, Leipzig, 1928.
6. W. F. Osgood, *Lehrbuch der Funktionentheorie*, vol. II, part 1, 2d edition, Leipzig, 1929.
7. C. Riquier, *La Méthode des Fonctions Majorantes et les Systèmes d'Equations aux Dérivées Partielles*, Mémorial des Sciences Mathématiques, fasc. 32, Paris, 1928.
8. C. Riquier, *Les Systèmes d'Equations aux Dérivées Partielles*, Paris, 1910.
9. J. F. Ritt, *Differential Equations from the Algebraic Standpoint*, New York, 1932.
10. J. M. Thomas, *The condition for an orthonomic differential system*, Transactions of the American Mathematical Society, vol. 34(1932), pp. 332–338.
11. J. M. Thomas, *Differential Systems*, New York, 1937.
12. J. M. Thomas, *Positive solutions of binomial inequalities*, this Journal, vol. 7(1940), pp. 291–297.
13. J. M. Thomas, *Riquier's existence theorems*, Annals of Mathematics, (2), vol. 30(1929), pp. 285–310.
14. J. M. Thomas, *Riquier's existence theorems*, Annals of Mathematics, (2), vol. 35(1934), pp. 306–311.

Duke University.

POSITIVE SOLUTIONS OF BINOMIAL INEQUALITIES

By Joseph Miller Thomas

1. Introduction. Let S be a finite set of inequalities of the form

$$m > p, \tag{1.1}$$

where m, p are monomials in the unknowns $x_1, x_2, \cdots, x_n$ with positive coefficients. We give a method of testing S for consistency and solving it in the set of all positive numbers. Although the sign $>$ is employed throughout the discussion of §§1–5, the method is valid for an arbitrary distribution of $>$, $\geqq$, $=$, except that the restriction to $>$ is essential in Theorem 4.1, as it is in §6. When various signs occur, they must be multiplied in accordance with the following table.

	$>$	$\geqq$	$=$
$>$	$>$	$>$	$>$
$\geqq$	$>$	$\geqq$	$\geqq$
$=$	$>$	$\geqq$	$=$

In other words, although $\geqq$ and $>$ when multiplied imply both $\geqq$ and $>$, the stronger inequality is defined as the product.

The order of the terms in an equality $m = p$ is not determined. When $m = p$ is present, both $m = p$ and $p = m$ must be included, if elimination is to proceed by multiplication alone. The method is, of course, valid for systems composed entirely of equations.

The method is also applicable when the exponents, instead of being nonnegative integers, are any real numbers. As a consequence, solving any finite set of linear inequalities in the real field is equivalent to solving a system (1.1) in the positive numbers. From this standpoint the method can be identified with the process of elimination developed by Dines.[1]

2. Reduction processes. If S contains in addition to (1.1) the inequality

$$q > t, \tag{2.1}$$

then S implies

$$(m - p)(q - t) > 0,$$

Received May 2, 1940.

[1] L. L. Dines, *Systems of linear inequalities*, Annals of Mathematics, (2), vol. 20(1918–1919), pp. 191–199.

or

$$mq - pt - t(m - p) - p(q - t) > 0.$$

Since $-t(m - p) - p(q - t)$ is negative for every solution of S, we must have

$$mq > pt \tag{2.2}$$

for every solution of S. Accordingly, we may multiply inequalities member for member. We shall call (2.2) the *product* of (1.1) and (2.1).

Similarly, if q is a monomial with positive coefficient and (1.1) implies

$$mq > pq.$$

The converse is also true.

By repeated multiplication from (1.1) we deduce

$$m^k > p^k. \tag{2.3}$$

Hence any inequality in S can be replaced by its k-th power without altering the solution of S. Moreover, (2.3) is seen to be equivalent to $m > p$ because the quotient $(m^k - p^k)/(m - p)$ is a polynomial in m, p with positive coefficients.

By the highest common factor (m, p) of m, p we shall understand the monomial which divides both m and p, which has maximum possible degree and which has coefficient $+1$. Without loss of generality we may suppose that for every inequality (1.1) of S

$$(m, p) = 1.$$

The two inequalities (1.1) and (2.1) will be called *relatively prime* if

$$(m, t)(p, q) = 1.$$

This condition is, of course, equivalent to the two $(m, t) = (p, q) = 1$.

From the foregoing discussion it should be clear that the following principles can be used to augment the system S or to modify it without altering its solution.

(i) $m > p$ implies $m^k > p^k$ and conversely.

(ii) $m > p$ and $q > t$ imply $\dfrac{mq}{(m, t)(p, q)} > \dfrac{pt}{(m, t)(p, q)}$.

If $(m, p) = (q, t) = 1$, the members of the last inequality in (ii) are relatively prime.

3. **Solution of the system.** To solve a system S, separate its inequalities into two sets S_1, T_1, putting into S_1 all those inequalities involving the first unknown, which for convenience we temporarily write without its index as x, and into T_1 all those free of x. The members of S_1 may be of two types:

$$x^\alpha m > p, \qquad q > x^\beta t, \tag{3.1}$$

where m, p, q, t do not involve x.

If S_1 contains only members of the first type, x must satisfy a set of inequalities of the form

$$x > \left(\frac{p}{m}\right)^{1/\alpha},$$

which are obviously consistent. A similar conclusion is reached if S_1 contains only members of the second type.

If S_1 contains members of both types, x must satisfy a set of inequalities of the form

$$\left(\frac{q}{t}\right)^{1/\beta} > x > \left(\frac{p}{m}\right)^{1/\alpha}. \tag{3.2}$$

In order that (3.2) be consistent, we must have

$$\left(\frac{q}{t}\right)^{1/\beta} > \left(\frac{p}{m}\right)^{1/\alpha}$$

a condition equivalent to

$$m^\beta q^\alpha > p^\beta t^\alpha. \tag{3.3}$$

Conversely, when (3.3) is satisfied, (3.2) defines a segment on the x-axis on which x must lie. It is convenient to think of the positive direction on the axis as leading from right to left.

The conditions (3.2) are to be formed for every pair of inequalities of different types in S_1. The corresponding conditions (3.3) are to be adjoined to T_1 to give a system S' in $x_2, \cdots, x_n$. For any solution of S' the conditions (3.2) define segments. Since each left endpoint occurs in combination with every right endpoint, each left endpoint is to the left of every right endpoint. Hence all the segments have in common a segment leading from the least left member to the greatest right member, and any point on this segment will serve for x.

Next S' is to be separated into two sets S_2, T_2 with respect to the second unknown and T_2 is to be augmented to give S''.

Repetition gives a sequence $S_1, S_2, \cdots, S_n, S_{n+1}$, where S_i $(i = 1, 2, \cdots, n)$ involves only unknowns of index at least equal to i and S_{n+1} involves only constants. The condition for consistency is that S_{n+1} *contain only valid inequalities.*

We note that a result equivalent to (3.3) is obtained by applying the operations (i), (ii) of §2 to (3.1).

4. **Equivalence to linear systems.** Although in the preceding discussion the use of the terms "monomial" and "highest common factor" has supposed that

$$m = ay_1^{i_1}y_2^{i_2} \cdots y_n^{i_n},$$

where the i's are non-negative integers, all of the operations, except the inconsequential operation of removing the highest common factor, are valid if the i's are any non-negative numbers.

For our next purpose it is convenient to write (1.1) as

$$cy_1^{a_1}y_2^{a_2} \cdots y_n^{a_n} > 1. \tag{4.1}$$

Since $\log x$ is real and increasing for all positive x, inequality (4.1) is equivalent to

$$a_1 \log y_1 + a_2 \log y_2 + \cdots + a_n \log y_n + \log c > 0. \tag{4.2}$$

Hence the linear inequality

$$a_1x_1 + a_2x_2 + \cdots + a_nx_n + b > 0 \tag{4.3}$$

is solved by finding the positive solutions of (4.1) in which $c = e^b$ and then putting $x_i = \log y_i$.

It is immediately seen that the operations we have applied to inequalities having the form (4.1) when translated into operations on inequalities having the form (4.3) become the operations of elimination employed by Dines (loc. cit.).

In illustration we set

$$x_3 = \log x, \qquad x_4 = \log y, \qquad x_1 = \log z, \qquad x_2 = \log t$$

in Dines' example on p. 198. The left members of the inequalities of form (4.1) are

$$xy^{-1}zt^{-3}, \qquad xy^2z^2t^{-3}, \qquad x^3yz^3t^{-2}, \qquad x^{-1}y^2z^{-2}t^2.$$

Solving, we find

$$y^2z^{-2}t^2 > x > yz^{-1}t^3, \qquad y^{-2}z^{-2}t^3, \qquad y^{-1/3}z^{-1}t^{2/3}$$
$$y > zt, \qquad t^{1/4}, \qquad z^{3/7}t^{-4/7}.$$

This result is easily identified with the second solution given on p. 198 of Dines' article.

A result useful in applications is

THEOREM 4.1. *A consistent system of linear inequalities, all of which have the sign $>$, has a rational solution. If the inequalities are homogeneous, there is an integral solution.*

Let x_i be a set of real numbers satisfying a system of linear inequalities like (4.3). Let $\lim_{t \to \infty} x_{it} = x_i$, where x_{it} is a sequence of rational numbers. Then

$$\lim_{t \to \infty} \left(\sum a_i x_{it} + b\right) > 0.$$

Hence there is an N such that

$$\sum a_i x_{it} + b > 0$$

provided $t > N$. For each of the other inequalities there is a similar N. If t is taken greater than the largest of them, a rational solution of the system is found.

If x_i is a rational solution of a homogeneous system and c is the positive least common multiple of the denominators of the x's, then cx_i is a solution in integers. The example $x + 1 > 0$, $-x > 0$ shows that a consistent non-homogeneous system does not necessarily have a solution in integers.

5. The homogeneous subset. Let the inequalities of the system S considered in §2 be separated into three sets in the following way. Every homogeneous inequality, that is, one whose left and right members have the same degree, is placed in a set H. A non-homogeneous inequality is placed in the set L or R according as its left or right member has higher degree.

The fundamental operation (i) applied to an inequality of any one of the three categories gives an inequality to be placed in the same category. Similarly, the product of two inequalities falls in a category determined by the following scheme:

$$HH = H, \qquad HL = L, \qquad HR = R, \qquad LL = L, \qquad RR = R.$$

Only when a member of L is multiplied by a member of R may the category of the product vary.

Now suppose R is vacuous. The combination of H with L or of L with L gives only inequalities of the category L. Since the inequalities of S_{n+1} are homogeneous, the set S_{n+1} coincides with the corresponding set for H. Hence we have

THEOREM 5.1. *When either R or L is vacuous, S is consistent if and only if the homogeneous set H is.*

This result can be verified directly, and at the same time the solution of S thrown back upon the solution of H, in the following manner. We suppose the set R vacuous and let S contain the inequality (1.1), where the degree of m is k more than the degree of p. Put $x_i = \lambda\xi_i$. Inequality (1.1) can then be written

$$\lambda > \left(\frac{\pi}{\mu}\right)^{1/k}, \tag{5.1}$$

where μ, π arise from m, p by replacing each x by the corresponding ξ.

Let the ξ's be a solution of H. For such ξ's determine a λ satisfying all conditions like (5.1) arising from L. Then $x_i = \lambda\xi_i$ satisfy both H and L, that is, S.

The labor of solution may be greatly diminished by this device.

6. The case of coefficients equal to unity. In this section the signs must all be $>$. Suppose the coefficients of all monomials in S are unity. Then all coefficients in $S_1, \cdots, S_{n+1}$ are also unity. Since the monomials of S_{n+1} are of degree zero, the only possible member for S_{n+1} is $1 > 1$, and S is consistent or not according as S_{n+1} is vacuous or not.

For a consistent S with unit coefficients let a system S' be formed by replacing each inequality $m > p$ by

$$m > cp,$$

where c is an arbitrary positive constant which may vary from one inequality to another. S'_{n+1} consists of inequalities among products of the c's. If it were not vacuous, placing the c's equal to unity would give the inequality $1 > 1$ which would have to be in S_{n+1}. Hence S' is consistent.

Since $\log c$ in (4.2) is zero if and only if $c = 1$, the systems considered in this section have for correspondents the homogeneous linear systems. The result just established means that if the homogeneous linear system is consistent, the corresponding non-homogeneous system is consistent for arbitrary values of the constant terms, as is also easily seen directly.

Moreover, the dual of the second part of Theorem 4.1 is

THEOREM 6.1. *If a system of binomial inequalities in which every sign is $>$ and every coefficient is unity has a solution in positive numbers, it has a solution in which each unknown is an integral power of any arbitrarily fixed real number greater than unity. In particular, it has a solution in integers.*

7. **A particular linear system.** The system

$$\begin{aligned} & t_i \geqq 0, \\ & p_i + t_i \geqq q_{ij} + t_j \qquad (i, j = 1, 2, \cdots, r), \end{aligned} \tag{7.1}$$

where the p's and q's are known non-negative integers and (non-negative) integral values are to be determined for the t's, is of interest in the study of systems of partial differential equations. It is convenient to denote by (i, j) the inequality on the second line of (7.1) and by $(i, j) + (k, l)$ the result of adding two inequalities member for member.

Rather than convert (7.1) into a binomial system, we apply the elimination process directly to (7.1). Since the coefficients of the t's are equal to unity, this process amounts to adding an inequality with a given t on the left to one with the same t on the right.

If $t_i \geqq 0$ is added to (j, i), there results an inequality with a t in its left member. Some t will remain there no matter how many inequalities of the type (k, l) are added to it. Hence in forming the conditions of consistency the first line of (7.1) is to be ignored. This of course also follows from Theorem 5.1.

If we make $j = i$ in (7.1) we get

$$p_i \geqq q_{ii}. \tag{7.2}$$

If $l > 1$ and if $i_1, i_2, \cdots, i_l$ are any l distinct integers from the range $1, 2, \cdots, r$, then

$$(i_1, i_2) + (i_2, i_3) + \cdots + (i_{l-1}, i_l) + (i_l, i_1) \tag{7.3}$$

is free of the t's and hence is a condition of consistency. Moreover, the totality of conditions (7.3) for all possible distinct cycles $i_1i_2 \cdots i_l$, taken together with (7.2) which correspond to the cycles of length one, contains all the conditions of consistency. To see this, we remark that if (i_1, i_2), (i_2, i_3) are added,

there results an inequality with right member containing t_{i_3}, which must subsequently be eliminated by adding an inequality (i_3, i_4). And so on. The last inequality added must eliminate both the remaining t's.

Consequently we have

THEOREM 7.1. *The system* (7.1) *is consistent in the ring of integers if and only if*

$$\sum p_i \geqq \sum q_{ij}, \tag{7.4}$$

where i runs through all distinct cycles of length from 1 *to r and j has the value immediately following that of i in the cycle.*

DUKE UNIVERSITY.

INVARIANTS OF SYSTEMS OF SECOND ORDER LINEAR DIFFERENTIAL EQUATIONS

By Jack Levine

1. **Introduction.** As a first step in the discussion of ruled surfaces, Wilczynski[1] obtained the seminvariants and invariants of a set of two differential equations in two dependent variables of the form (2.1). Recently, Barnett and Reingold[2] have treated a similar problem for n equations in n dependent variables, and obtained sets of functionally independent seminvariants of order 1 for $n = 2$, 3, and 4, and an independent set of order 0 for general[3] n.

In this paper we consider the problem of obtaining seminvariants, invariants, and covariants of a general order r of the system (2.1) containing n equations in n dependent variables. The methods of tensor analysis are used as these are suggested by the nature of the problem.

In §3 it is shown that every seminvariant is a function of quantities P^i_j (defined in §2) and their covariant derivatives with components $P^i_{j;\alpha}$. By using the components $P^i_{j;\alpha}$ as independent variables, the number of equations in the complete system defining the seminvariants remains fixed for all orders r and a given n. In the previous treatments the number of such equations increased with r.

In §7 a functionally independent set of seminvariants of general order r (>0) is obtained for $n = 2$ and $n = 3$, and in §8, it is shown that the tensor seminvariants with components $P^i_{j;\alpha}$ form a complete set of such invariants and can be thus used in the equivalence problem of two systems (2.1).

2. **Seminvariants.** We shall consider the system of differential equations[4]

$$\frac{d^2y^i}{dx^2} + L^i_j(x)\,\frac{dy^j}{dx} + M^i_j(x)y^j = 0, \tag{2.1}$$

where L^i_j and M^i_j are arbitrary functions of x possessing as many derivatives as are necessary for the purpose of our discussion. It is shown in W that the most general transformation of the y which changes (2.1) into a similar form is

$$y^i = a^i_j(x)\bar{y}^j, \tag{2.2}$$

where $|\, a^i_j \,| \not\equiv 0$, but the a^i_j are otherwise arbitrary.

Received May 3, 1940.

[1] E. J. Wilczynski, *Projective Differential Geometry of Curves and Ruled Surfaces*, Chapter IV. We call this reference W.

[2] I. A. Barnett and H. Reingold, *Invariants of a system of linear homogeneous differential equations of the second order*, this Journal, vol. 6(1940), pp. 141–147. We refer to this paper as BR.

[3] See §2 of this paper for the definition of the order of a seminvariant.

[4] All indices throughout this paper have the range 1, 2, 3, $\cdots$, n, unless otherwise stated. Also, we assume $n \geq 2$ throughout.

If we replace y^i in (2.1) by means of (2.2), we obtain the transformed system

$$\frac{d^2\bar{y}^i}{dx^2} + \bar{L}^i_j(x)\frac{d\bar{y}^j}{dx} + \bar{M}^i_j(x)\bar{y}^j = 0, \tag{2.3}$$

where

$$\bar{L}^i_j = \bar{a}^i_h\left(L^h_k a^k_j + 2\frac{da^h_j}{dx}\right), \tag{2.4}$$

$$\bar{M}^i_j = \bar{a}^i_h\left(M^h_k a^k_j + L^h_k\frac{da^k_j}{dx} + \frac{d^2a^h_j}{dx^2}\right), \tag{2.5}$$

the $\bar{a}^i_j$ being defined by $\bar{a}^i_j a^j_k = \delta^i_k$, $\bar{a}^i_j a^k_i = \delta^k_j$.

An (absolute) seminvariant of order r of (2.1) is defined as a function

$$f\left(L^i_j, \frac{dL^i_j}{dx}, M^i_j; \frac{d^2L^i_j}{dx^2}, \frac{dM^i_j}{dx}; \cdots; \frac{d^{r+1}L^i_j}{dx^{r+1}}, \frac{d^rM^i_j}{dx^r}\right), \tag{2.6}$$

which transforms under (2.2) by

$$f\left(\bar{L}^i_j, \cdots, \frac{d^r\bar{M}^i_j}{dx^r}\right) = f\left(L^i_j, \cdots, \frac{d^rM^i_j}{dx^r}\right), \tag{2.7}$$

i.e., f retains its form as a function of the L's and M's under (2.2).

If (2.4) is differentiated with respect to x and $d^2a^h_j/dx^2$ is eliminated by means of (2.5), we obtain

$$\bar{P}^i_j = P^h_k a^k_j \bar{a}^i_h, \tag{2.8}$$

where

$$P^i_j \equiv \frac{dL^i_j}{dx} + \tfrac{1}{2}L^i_h L^h_j - 2M^i_j, \tag{2.9}$$

and $\bar{P}^i_j$ are defined in a similar manner in terms of $\bar{L}^i_j$, $\bar{M}^i_j$. We thus see that P^i_j are the components of a mixed tensor under the transformation (2.2).

From the relations (2.8) we obtain a new tensor by differentiation with respect to x and elimination of da^i_h/dx by (2.4). This leads to

$$\bar{P}^i_{j;1} = P^h_{k;1} a^k_j \bar{a}^i_h, \tag{2.10}$$

with

$$P^i_{j;1} \equiv \frac{dP^i_j}{dx} + \tfrac{1}{2}(P^h_j L^i_h - P^i_h L^h_j). \tag{2.11}$$

We call the tensor with components $P^i_{j;1}$ the first covariant derivative of P^i_j, and the L^i_j can be thought of as the coefficients of a linear connection. In general, the covariant derivative of order α of P^i_j is defined by[5]

$$P^i_{j;\alpha} = \frac{dP^i_{j;\alpha-1}}{dx} + \tfrac{1}{2}(P^h_{j;\alpha-1}L^i_h - P^i_{h;\alpha-1}L^h_j). \tag{2.12}$$

[5] The components P^i_j, $P^i_{j;1}$ correspond to G_{ij}, U_{ij} of BR.

3. **A replacement theorem for seminvariants.** It is possible to find a set of coördinates $\bar{y}^i$ in which the quantities $\bar{L}^i_j = 0$. From (2.4) we see it is sufficient to make

$$\frac{da^h_j}{dx} = -\tfrac{1}{2}L^h_k a^k_j. \tag{3.1}$$

These equations are of the same form for all values of j and hence possess n independent sets of solutions[6] $a^1_j, a^2_j, \cdots, a^n_j$.

The equations (2.1) with $L^i_j = 0$ will be said to be in semi-canonical form, and thus every such system can be so reduced.

Consider now any seminvariant (2.6). By changing to a semi-canonical form, f becomes

$$f = f\left(0, 0, \bar{M}^i_j; 0, \frac{d\bar{M}^i_j}{dx}; \cdots; 0, \frac{d^r\bar{M}^i_j}{dx^r}\right). \tag{3.2}$$

But from (2.9) and (2.12) in semi-canonical form we see that

$$\bar{M}^i_j = -\tfrac{1}{2}\bar{P}^i_j, \qquad \frac{d^\alpha \bar{M}^i_j}{dx^\alpha} = -\tfrac{1}{2}\bar{P}^i_{j;\alpha}.$$

Hence f of (3.2) reduces to a function of $\bar{P}^i_j$ and its first r covariant derivatives, but since a seminvariant must retain its form under any transformation (2.2), we can state the following replacement theorem.

THEOREM 3.1. *Every seminvariant* (2.6) *can be expressed as a function of* P^i_j *and its covariant derivatives by replacing* L^i_j *and its derivatives by zero and* $d^\alpha M^i_j/dx^\alpha$ *by* $-\tfrac{1}{2}P^i_{j;\alpha}$.

We can thus assume a seminvariant of order r to be written in the form

$$f(P) \equiv f(P^i_j, P^i_{j;1}, \cdots, P^i_{j;r}). \tag{3.3}$$

4. **The complete system of differential equations.** To obtain the system of partial differential equations satisfied by $f(P)$ we use the infinitesimal transformation corresponding to (2.2),

$$y^i = \bar{y}^i + \epsilon\phi^i_j(x)\bar{y}^j. \tag{4.1}$$

Here

$$a^i_j = \delta^i_j + \epsilon\phi^i_j, \qquad \bar{a}^i_j = \delta^i_j - \epsilon\phi^i_j, \tag{4.2}$$

and substituting these values in the transformation equations of $P^i_{j;\alpha}$, we find

$$\bar{P}^i_{j;\alpha} = P^i_{j;\alpha} + \epsilon(P^i_{m;\alpha}\phi^m_j - P^m_{j;\alpha}\phi^i_m), \tag{4.3}$$

so that

$$\left(\frac{d\bar{P}^i_{j;\alpha}}{d\epsilon}\right)_{\epsilon=0} = P^i_{m;\alpha}\phi^m_j - P^m_{j;\alpha}\phi^i_m. \tag{4.4}$$

[6] E. Goursat, *Mathematical Analysis*, vol. 2, part 2, p. 50. We refer to this as G. For the corresponding result for $n = 2$ see W, p. 114.

If $f(\bar{P})$ be expanded about $\epsilon = 0$, we obtain from (4.4) and the relation $f(\bar{P}) = f(P)$,

$$\left[\frac{\partial f(P)}{\partial P^i_{j;\alpha}} (P^i_{m;\alpha}\delta^k_j - P^k_{j;\alpha}\delta^i_m)\right]\phi^m_k = 0.$$

Since the ϕ^m_k are arbitrary, we have

$$X^k_m(r)f \equiv \sum_{\alpha=0}^{r} \begin{pmatrix} i & k & \\ j & m & \alpha \end{pmatrix} \frac{\partial f}{\partial P^i_{j;\alpha}} = 0, \tag{4.5}$$

as the system of equations satisfied by $f(P)$.[7]

In (4.5) the quantities $\begin{pmatrix} i & k & \\ j & m & \alpha \end{pmatrix}$ are defined by

$$\begin{pmatrix} i & k & \\ j & m & \alpha \end{pmatrix} = P^i_{m;\alpha}\delta^k_j - P^k_{j;\alpha}\delta^i_m. \tag{4.6}$$

It is easy to show by direct calculation that $(X^k_m, X^i_j)f = \delta^i_m X^k_j f - \delta^k_j X^i_m f$, and this proves that (4.5) form a complete system.

We shall say that the system (4.5) is of rank R if it contains exactly R independent equations. Since from (4.5) and (4.6) $X^i_i(r)f \equiv 0$, we see that (4.5) is of rank $n^2 - 1$ at most. From the form of (4.5), if $X^k_m(r)f = 0$ are of rank $n^2 - 1$, then so also are $X^k_m(r + 1)f = 0$.

We now show that if $X^k_m(1)f = 0$ are of rank $n^2 - 1$ for n variables (y^i), they are of rank $(n + 1)^2 - 1$ for $n + 1$ variables (y^i, y^N), $N = n + 1$.

For $r = 1$, (4.5) reduce to

$$\begin{pmatrix} i & k \\ j & m \end{pmatrix} \frac{\partial f}{\partial P^i_j} + \begin{pmatrix} i & k & \\ j & m & 1 \end{pmatrix} \frac{\partial f}{\partial P^i_{j;1}} = 0. \tag{4.7}$$

The matrix of $2n^2$ columns and $n^2 - 1$ rows

$$\left\| \begin{pmatrix} i & k \\ j & m \end{pmatrix} \quad \begin{pmatrix} i & k & \\ j & m & 1 \end{pmatrix} \right\|, \tag{4.8}$$

in which i, j represent columns and k, m rows, with $k = m = 1$ omitted, is assumed of rank $n^2 - 1$, and hence contains a non-zero determinant, D_n, of order $n^2 - 1$.

In the matrix (4.8) for N variables, select the $n^2 - 1$ columns containing D_n, the $2n$ columns of the coefficients of $\partial f/\partial P^i_N$ and $\partial f/\partial P^N_i$, and the column of the coefficients of $\partial f/\partial P^N_{1;1}$. If in the resulting determinant D_{n+1} of order $N^2 - 1$ we put $P^i_N = P^N_i = P^N_{2;1} = P^N_{3;1} = \cdots = P^N_{n;1} = 0$, it will reduce to

$$D_{n+1} = \begin{vmatrix} D_n & 0 & 0 & 0 \\ * & 0 & -D^i_m & * \\ * & D^i_m & 0 & 0 \\ 0 & 0 & 0 & -P^N_{1;1} \end{vmatrix} = \pm D_n P^N_{1;1} \,|\, D^i_m|^2 \not\equiv 0. \tag{4.9}$$

[7] For a corresponding treatment of affine differential invariants see T. Y. Thomas and A. D. Michal, *Differential invariants of affinely connected manifolds*, Annals of Math., vol. 28(1927), pp. 196–236. We refer to this as TM.

In (4.9), $D^i_m = P^i_m - \delta^i_m P^N_N$, and the *'s represent terms whose values are not needed.

Now if $n = 2$, the rank of (4.7) is easily shown to be $2^2 - 1 = 3$, and hence we can state the following theorem.

THEOREM 4.1. *The complete system $X^k_m(r)f = 0$ for the determination of seminvariants of order r in n variables contains $n^2 - 1$ independent equations for $r \geqq 1$.*[8]

It is shown in BR that $X^i_j(0)f = 0$ is of rank $n^2 - n$. This can also be seen by considering the determinant obtained from D_{n+1} above by omitting the last row and column, where now D_n represents a non-zero determinant of order $n^2 - n$ for the n variable case. From the fact that trace $(P^k X) = 0$ $(k = 0, 1, \cdots, n - 1)$, where P, X represent the matrices $\| P^i_j \|$, $\| X^i_j \|$, the rank of $X^i_j(0)f = 0$ is at most $n^2 - n$, and as for $n = 2$ the rank can be shown to be $2 = 2^2 - 2$, the rank is $n^2 - n$ for general n.

Since there are $(r + 1)n^2$ independent variables in the complete system (4.5), we can now state

THEOREM 4.2. *The number of (absolute) seminvariants in a fundamental set*[9] *of order r for the system* (2.1) *in n dependent variables is given by*

$$N(n, r) = (r + 1)n^2 - (n^2 - 1) = rn^2 + 1 \ (r \geqq 1),$$
$$N(n, 0) = n^2 - (n^2 - n) = n.$$

5. **Semi-covariants.** An (absolute) semi-covariant is a function of the form $f(y^i, dy^i/dx, L^i_j, \cdots)$ which transforms under (2.2) by

$$f\left(\bar{y}^i, \frac{d\bar{y}^i}{dx}, \bar{L}^i_j, \cdots\right) = f\left(y^i, \frac{dy^i}{dx}, L^i_j, \cdots\right). \tag{5.1}$$

It is not necessary to assume the presence of higher order derivatives of y as they can be eliminated by use of (2.1).

If (2.2) is differentiated with respect to x and da^i_j/dx is eliminated by means of (2.5), we obtain

$$\bar{Y}^h = Y^i a^h_i, \tag{5.2}$$

where

$$Y^i \equiv \frac{dy^i}{dx} + \tfrac{1}{2} y^h L^i_h; \tag{5.3}$$

so Y^i is the covariant derivative of y^i.

If we transform y^i to a semi-canonical form $\bar{y}^i$, then $\bar{Y}^i = d\bar{y}^i/dx$, and (5.1) shows every semi-covariant can be expressed in the form

$$f(y^i, Y^i; P^i_j, \cdots, P^i_{j;r}). \tag{5.4}$$

[8] A theorem similar to this for $r = 1$ is proved in BR in a different manner, but we have given the present proof as it is used in later sections.

[9] For the definition of a fundamental set see TM, p. 216.

From (4.1), (4.2), and (5.2), we obtain

$$\left(\frac{d\bar{y}^i}{d\epsilon}\right)_0 = -\phi^i_h y^h, \qquad \left(\frac{d\bar{Y}^i}{d\epsilon}\right)_0 = -\phi^i_h Y^h,$$

and proceeding in the same way used to obtain (4.5), we find

$$S^k_m(r, s)f \equiv \sum_{\alpha=0}^{r} \begin{pmatrix} i & k & \\ j & m & \alpha \end{pmatrix} \frac{\partial f}{\partial P^i_{j;\alpha}} - y^k \frac{\partial f}{\partial y^m} - Y^k \frac{\partial f}{\partial Y^m} = 0 \tag{5.5}$$

as the complete system satisfied by (5.4), which we shall say is of order (r, s), where $s = 0$ if f depends on y^i but not Y^i, and $s = 1$ if f contains both y^i, Y^i.

The n^2 equations are independent for all values of n, r, s. To see this for $r > 0$, we select the $n^2 - 1$ columns of (4.8) (with the row $k = m = 1$ included), containing D_n and to them add the column of the coefficients of $\partial f/\partial y^1$ obtained from (5.5). If we put $y^2 = y^3 = \cdots = y^n = 0$ in this last column, the resulting determinant has the value $y^1 D_n$. This shows (5.5) are independent for $r > 0$.

If $r = 0$, assume (5.5) independent for n variables, and form D_{n+1} as in (4.9), where now D_n is a non-zero determinant of order n^2 corresponding to n variables. Also, in place of the last column of (4.9) we use the column of the coefficients of $\partial f/\partial y^N$ in which we put $y^i = 0$. The resulting determinant of order $(n + 1)^2$ has the value $y^N D_n \mid D^i_m \mid^2 \not\equiv 0$. For $n = 2$, $r = 0$, the four equations (5.5) are independent, and so this is true for all n.

There being $(r + 1)n^2 + (s + 1)n$ independent variables and n^2 independent equations in the complete system (5.5) we have

THEOREM 5.1. *The number of (absolute) semi-covariants in a fundamental set of order (r, s) for the system (2.1) in n dependent variables is given by*

$$N(n, r, s) = rn^2 + (s + 1)n \qquad (r \geqq 0; s = 0, 1).$$

6. **Invariants.** An (absolute) invariant is a seminvariant which remains unchanged in form under an arbitrary transformation of the independent variable x. If we take this transformation in the form

$$\bar{x} = \bar{x}(x), \qquad \Delta = \frac{d\bar{x}}{dx}, \tag{6.1}$$

then we find

$$\bar{L}^i_j(\bar{x}) = \frac{1}{\Delta}(L^i_j(x) + \delta^i_j \eta), \qquad \bar{M}^i_j(\bar{x}) = \frac{1}{\Delta^2} M^i_j(x), \tag{6.2}$$

$$\bar{P}^i_k(\bar{x}) = \frac{1}{\Delta^2}(P^i_k + \delta^i_k \mu), \tag{6.3}$$

where

$$\eta = \frac{\Delta'}{\Delta}, \qquad \mu = \eta' - \tfrac{1}{2}\eta^2 = \frac{\Delta''}{\Delta} - \frac{3}{2}\left(\frac{\Delta'}{\Delta}\right)^2, \tag{6.4}$$

and primes refer to derivatives with respect to x.

Under the infinitesimal form of (6.1),

$$\bar{x} = x + \epsilon\phi(x), \tag{6.5}$$

we have

$$\Delta = 1 + \epsilon\phi', \qquad \frac{1}{\Delta^\omega} = 1 - \epsilon\omega\phi', \qquad \mu = \epsilon\phi^{(3)}, \tag{6.6}$$

$$\bar{P}^i_k = P^i_k + \epsilon(\delta^i_k\phi^{(3)} - 2P^i_k\phi'), \tag{6.7}$$

$$\bar{P}^i_{k;1} = P^i_{k;1} + \epsilon(\delta^i_k\phi^{(4)} - 2P^i_k\phi^{(2)} - 3\phi' P^i_{k;1}), \tag{6.8}$$

and in general, if we express the transformation of $P^i_{k;\alpha}$ in the form

$$\bar{P}^i_{k;\alpha} = P^i_{k;\alpha} + \epsilon P^i_{k\alpha\beta}\phi^{(\beta)}, \tag{6.9}$$

then we find

$$\bar{P}^i_{k;\alpha+1} = P^i_{k;\alpha+1} + \epsilon(P^i_{k\alpha\beta}\phi^{(\beta+1)} + P^i_{k\alpha\beta;1}\phi^{(\beta)} - P^i_{k;\alpha+1}\phi'). \tag{6.10}$$

In (6.9), for each value of α, the index β takes on the values $1, 2, 3, \cdots, \alpha + 3$. If we carry through for the seminvariant (3.3) a process similar to that used to obtain the system (4.5), but use the relations (6.5)–(6.9), we obtain the following set of equations satisfied by an invariant of order r (in addition to (4.5))

$$X_\beta(r)f \equiv \sum_{\alpha=0}^{r} P^i_{j\alpha\beta}\frac{\partial f}{\partial P^i_{j;\alpha}} = 0 \qquad (\beta = 1, 2, \cdots, r+3). \tag{6.11}$$

We shall call the combined system (4.5), (6.11) the equations E_{rn} of order r. The equations $E_{r+1,n}$ are obtained by adding terms to the equations (6.11), for $\beta = 1, 2, \cdots, r + 3$, and by adding one additional equation (to the set (6.11)). This new equation we call $F_{r+1,n}$, the final equation of order $r + 1$.

There are $n^2 + r + 3$ equations in E_{rn} $(r > 0)$, and their rank is $A(r, n) = n^2 + r + 2$ at most. It is easy to see that if E_{rn} is of rank $A(r, n)$ then $E_{r+1,n}$ is of rank $A(r + 1, n) = n + r + 3$. For $F_{r+1,n}$ has the form

$$\delta^i_j\frac{\partial f}{\partial P^i_{j;r+1}} = 0,$$

and a non-zero determinant A_{r+1} of order $A(r + 1, n)$ can be formed from the coefficients of $E_{r+1,n}$:

$$A_{r+1} = \begin{vmatrix} A_r & * \\ 0 & 1 \end{vmatrix},$$

where A_r represents a non-zero determinant of order $A(r, n)$ obtained from E_{rn} and the last column of A_{r+1} is the coefficients of $\partial f/\partial P^1_{1;r+1}$ in $E_{r+1,n}$ while the last row consists of A_r zeros and a 1.

We shall now show that if E_{1n} is of rank $A(1, n)$, then $E_{1,n+1}$ is of rank $A(1, n + 1)$. From (6.7) and (6.8) we find the system $E_{1,n+1}$ has the form

$$
\begin{aligned}
&X_1(1)f \equiv -2P_j^i \frac{\partial f}{\partial P_j^i} - 3P_{j;1}^i \frac{\partial f}{\partial P_{j;1}^i} = 0,\\
&X_2(1)f \equiv \qquad\qquad -2P_j^i \frac{\partial f}{\partial P_{j;1}^i} = 0,\\
&X_3(1)f \equiv \delta_j^i \frac{\partial f}{\partial P_j^i} \qquad\qquad = 0,\\
&X_4(1)f \equiv \qquad\qquad \delta_j^i \frac{\partial f}{\partial P_{j;1}^i} = 0,
\end{aligned}
\tag{6.12}
$$

where $i, j = 1, 2, \cdots, n, n + 1$.

Assume there exists a non-zero determinant A_{1n} of order $A(1, n)$ formed from the coefficients of E_{1n}. From $E_{1,n+1}$ we form a determinant $A_{1,n+1}$ of order $A(1, n + 1)$, its first $A(1, n)$ columns containing the columns of A_{1n}, and its last $2n + 1$ columns being formed in the same way as for D_{n+1} of (4.9).

If we put $P_N^i = P_i^N = P_{2;1}^N = \cdots = P_{n;1}^N = 0$ in $A_{1,n+1}$, then as seen from (6.12) all the elements in the last $2n + 1$ columns of the four rows corresponding to $X_\beta(1)f = 0$ will be zero except the element $-3P_{1;1}^N$ in the last column for $X_1(1)f = 0$. It is then easily shown that $A_{1,n+1}$ has the value $\pm A_{1n}P_{1;1}^N \mid D_m^i \mid^2 \neq 0$. Since for $n = 2$ it is not difficult to show E_{12} is of rank $A(1, 2)$, we have thus shown that E_{rn} is of rank $A(r, n)$ if $r \geqq 1$.

If $r = 0$, E_{0n} contains $A(0, n) = n^2 - n + 2$ equations which can be shown to be independent in the same manner used for E_{1n}.

Combining all these results gives us

THEOREM 6.1. *The number of (absolute) invariants in a fundamental system of order r of the system* (2.1) *is given by*

$$
\begin{aligned}
I(n, r) &= (n^2 - 1)r - 2 \qquad (r \geqq 1),\\
I(n, 0) &= n - 2.
\end{aligned}
$$

7. Independent sets of seminvariants. In BR, functionally independent sets of seminvariants of order 1 for $n = 2, 3$, and 4, and of order 0 for general n are obtained. We shall obtain such independent sets of general order r ($\geqq 1$) for $n = 2$ and $n = 3$.[10]

If we define seminvariant P_α, $P_{\alpha\beta}$ by

$$
P_\alpha \equiv P_{i;\alpha}^i, \qquad P_{\alpha\beta} \equiv P_{j;\alpha}^i P_{i;\beta}^j \quad (\alpha, \beta = 0, 1, 2, \cdots). \tag{7.1}
$$

[10] These are slightly different for $r = 1$ from those given in BR.

then the set of $4r + 1$ ($r > 0$) seminvariants

$$\text{(7.2)} \qquad \begin{cases} P_\alpha, \quad P_{\alpha\alpha} & (\alpha = 0, 1, \cdots, r), \\ P_{\alpha\beta} & (\alpha \neq \beta; \alpha = 0, 1; \beta = 1, 2, \cdots, r) \end{cases}$$

is functionally independent. (In (7.1), $\alpha = 0$ corresponds to P^i_j.)

We refer to the set (7.2) as $P_2(r)$. To prove $P_2(r)$ are independent, we assume $P_2(r-1)$ are, and hence possess a non-zero Jacobian determinant J_{r-1} of order $4r - 3$. $P_2(r)$ consists of $P_2(r-1)$ and four additional seminvariants P_r, P_{rr}, P_{0r}, P_{1r}. From the Jacobian matrix of $P_2(r)$ we select the determinant J_r of order $4r + 1$,

$$\text{(7.3)} \qquad J_r = \begin{vmatrix} J_{r-1} & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 1 \\ 0 & 2P^1_{1;r} & 2P^2_{1;r} & 2P^1_{2;r} & 2P^2_{2;r} \\ * & P^1_1 & P^2_1 & P^1_2 & P^2_2 \\ * & P^1_{1;1} & P^2_{1;1} & P^1_{2;1} & P^2_{2;1} \end{vmatrix}.$$

The last four rows of J_r correspond to the four added seminvariants in the order given, while its last four columns are the derivatives of these four with respect to $P^i_{j;r}$. From (7.3) we see $J_r = J_{r-1}\Delta_4$, where Δ_4 is a fourth order determinant which is not identically zero. It thus follows that if $P_2(r-1)$ are functionally independent, then so are $P_2(r)$. Now if $r = 1$, we easily show $P_2(1)$ are independent. Hence we have

THEOREM 7.1. *The set of* $4r + 1$ *seminvariants* (7.2) *is functionally independent for* $r \geqq 1$ *and* $n = 2$.

We consider now the case $n = 3$ ($r \geqq 1$), and shall show the set of $9r + 1$ seminvariants

$$\text{(7.4)} \qquad \begin{array}{llll} P_\alpha, & P_{\alpha\alpha}, & P_{\alpha\alpha\alpha} & (\alpha = 0, 1, \cdots, r), \\ P_{\alpha\beta}, & P_{\alpha\alpha\beta}, & P_{\alpha\beta\beta} & (\alpha \neq \beta; \alpha = 0, 1; \beta = 1, 2, \cdots, r), \\ P_{0011} & & & \end{array}$$

is functionally independent. Here,

$$P_{\alpha\beta\gamma} \equiv P^i_{j;\alpha} P^j_{k;\beta} P^k_{i;\gamma}, \qquad P_{0011} \equiv P^i_j P^j_k P^k_{m;1} P^m_{i;1}.$$

We call the set (7.4) $P_3(r)$, and let $F(r+1)$ represent the nine seminvariants added to $P_3(r)$ to form $P_3(r+1)$. We assume $P_3(r-1)$ are independent and possess a non-zero functional determinant K_{r-1}. For $P_3(r)$ we select the functional determinant of order $9r + 1$,

$$K_r = \begin{vmatrix} K_{r-1} & 0 \\ * & \dfrac{\partial F(r)}{\partial P^i_{j;r}} \end{vmatrix}.$$

To show $F_r = |\partial F(r)/\partial P^i_{j;r}| \not\equiv 0$, it is sufficient to put

$$P^i_j = 0 \quad (i \neq j), \qquad P^2_{1;1} = P^3_{1;1} = P^2_{1;r} = P^2_{3;r} = 0$$

in F_r, whence direct calculation shows $F_r \not\equiv 0$, and hence $K_r \not\equiv 0$. It remains to show $P_3(1)$ independent. From the Jacobian matrix of 10 functions and 18 independent variables P^i_j, $P^i_{j;1}$,

$$\left\| \frac{\partial P_3(1)}{\partial P^i_j}, \quad \frac{\partial P_3(1)}{\partial P^i_{j;1}} \right\|,$$

we select the ten columns formed from the derivatives with respect to

$$P^1_1, \quad P^1_2, \quad P^2_1, \quad P^1_{1;1}, \quad P^1_{2;1}, \quad P^2_{1;1}, \quad P^1_{3;1}, \quad P^2_{3;1}, \quad P^3_{1;1}, \quad P^3_{2;1}.$$

If we put $P^1_1 = P^2_{1;1} = P^1_{2;1} = P^2_{2;1} = P^3_{3;1} = P^3_{1;1} = P^3_{2;1} = 0$ in this determinant, it is not difficult to show it is not identically zero, and hence $P_3(1)$ are independent. This proves

THEOREM 7.2. *The set of* $9r + 1$ *seminvariants of* (7.4) *is functionally independent for* $r \geqq 1$ *and* $n = 3$.

8. **A complete set of tensor seminvariants.** Two systems of the form (2.1) are said to be equivalent if there exists a transformation of the form (2.2) which converts one system into the other. A set of seminvariants S is said to form a complete set if algebraic necessary and sufficient conditions for the equivalence of two systems (2.1) can be stated in terms of S.[11] In particular, S may consist of a set of tensor seminvariants. By such an invariant is meant a quantity with components $T^{ij\cdots k}_{lm\cdots p}(x)$ transforming under (2.2) by

$$\bar{T}^{ij\cdots k}_{lm\cdots p}(x) a^q_i a^r_j \cdots a^s_k = T^{qr\cdots s}_{tu\cdots v}(x) a^t_l a^u_m \cdots a^v_p.$$

The tensors with components $P^i_{j;\alpha}$ are examples of such seminvariants.

We shall prove that the totality of tensors with components $P^i_{j;\alpha}$ ($\alpha = 0, 1, 2, \cdots$) constitute a complete set of tensor seminvariants of the system (2.1).

We first prove a lemma which is a generalization of a theorem given in W, p. 115.

LEMMA. *The equations*

$$x = x(\bar{x}), \qquad y^i = h^i_j \bar{y}^j \bar{\Delta}^{\frac{1}{2}} \qquad \left(\bar{\Delta} = \frac{dx}{d\bar{x}}, \; |h^i_j| \neq 0\right) \tag{8.0}$$

give the most general transformations which leave the semi-canonical form of (2.1) *invariant,* $x(\bar{x})$ *being arbitrary, and* h^i_j *being constants.*

Under the simultaneous transformations (2.2), (6.1), a system (2.1) assumed in semi-canonical form, reduces to

$$\frac{d^2\bar{y}^i}{d\bar{x}^2} + \bar{L}^i_j(\bar{x})\frac{d\bar{y}^j}{d\bar{x}} + \bar{M}^i_j(\bar{x})\bar{y}^j = 0, \tag{8.1}$$

[11] See TM, p. 200 for a corresponding discussion for affine invariants.

where

$$\bar{L}^h_j = \frac{1}{\Delta^2}\left(2\Delta\bar{a}^h_i\frac{da^i_j}{dx} + \delta^h_j\frac{d\Delta}{dx}\right), \tag{8.2}$$

$$\bar{M}^h_j = \frac{1}{\Delta^2}\left(M^i_k a^k_j \bar{a}^h_i + \frac{d^2 a^i_j}{dx^2}\bar{a}^h_i\right). \tag{8.3}$$

If we wish (8.1) also to be in semi-canonical form, we have from (8.2), on placing $\bar{L}^h_j = 0$,

$$a^i_j(x) = \Delta^{-\frac{1}{2}}h^i_j, \qquad |h^i_j| \neq 0.$$

This proves the lemma.

From the form of (3.1) a unique set of solutions $a^i_j(x)$ is determined by the initial conditions[12]

$$x = q, \qquad a^i_j = h^i_j \qquad (q,\ h^i_j \text{ arbitrary consts., } |h^i_j| \neq 0). \tag{8.4}$$

Suppose now that (2.1) and

$$\frac{d^2\bar{y}^i}{dx^2} + \bar{L}^i_j(x)\frac{d\bar{y}^j}{dx} + \bar{M}^i_j(x)\bar{y}^j = 0 \tag{8.5}$$

are equivalent. A necessary condition evidently is that the sequence of equations

$$\begin{aligned} \bar{P}^i_j(x)a^k_i(x) &= P^k_i(x)a^i_j(x), \\ \bar{P}^i_{j;1}a^k_i &= P^k_{i;1}a^i_j\,, \\ &\cdots\cdots\cdots\cdots \end{aligned} \tag{8.6}$$

possess a numerical solution of the form (8.4). This condition is also sufficient as we now show.

Let the conditions (8.4) determine the semi-canonical transformations for (2.1) and (8.5) respectively:

$$y^i = a^i_j(x)z^j, \qquad \bar{y}^i = a^i_j(x)\bar{z}^j, \tag{8.7}$$

and let M^i_j, P^i_j be transformed under (8.7) to C^i_j, K^i_j, and $\bar{M}^i_j$, $\bar{P}^i_j$ to $\bar{C}^i_j$, $\bar{K}^i_j$. Also, we relate the coördinates z^i, $\bar{z}^i$ by

$$z^i = h^i_j\bar{z}^j. \tag{8.8}$$

From the relations

$$K^i_{j;\alpha}(x)a^k_i(x) = P^k_{i;\alpha}(x)a^i_j(x), \qquad \bar{K}^i_{j;\alpha}(x)a^k_i(x) = \bar{P}^k_{i;\alpha}(x)a^i_j(x), \tag{8.9}$$

and the expansions about $x = q$,

$$K^i_j(x) = \sum_{\alpha=0}^{\infty} K^i_{j;\alpha}(q)(x-q)^\alpha, \qquad \bar{K}^i_j(x) = \sum_{\alpha=0}^{\infty}\bar{K}^i_{j;\alpha}(q)(x-q)^\alpha, \tag{8.10}$$

[12] G, loc. cit.

we obtain by use of (8.4) and (8.6),

$$\bar{K}^i_j(x)h^k_i = K^k_i(x)h^i_j\,, \tag{8.11}$$

or

$$\bar{C}^i_j(x)h^k_i = C^k_i(x)h^i_j\,. \tag{8.12}$$

Thus, from (2.6), (8.8), (8.12), and the lemma with $x = \bar{x}$, we see that C^i_j, $\bar{C}^i_j$ are the forms of $M^i_j(x)$ in two related semi-canonical coördinate systems. From (8.7) and (8.8) we find

$$y^i = A^i_j(x)\bar{y}^j, \qquad A^i_j(x) = a^i_k(x)h^k_m\bar{a}^m_j(x), \qquad A^i_j(q) = h^i_j\,. \tag{8.13}$$

Hence from (8.12) and (8.13) it follows that (2.5) and (2.6) are satisfied with a^i_j replaced by A^i_j. We can thus state the following theorem.

THEOREM 8.1. *A necessary and sufficient condition that two systems of differential equations* (2.1) *and* (8.5) *be equivalent is that the infinite set of equations* (8.6) *possess a numerical solution* (8.4).

9. Some examples. It is easily shown that the quantities p_{ij}, q_{ij}, u_{ij}, v_{ij}, w_{ij} of W when expressed in our notation have the form ($n = 2$)

$$p_{ij} = L^i_j\,, \quad q_{ij} = M^i_j\,, \quad u_{ij} = 2P^i_j\,, \quad v_{ij} = 4P^i_{j;1}\,, \quad w_{ij} = 8P^i_{j;2}\,. \tag{9.1}$$

Also the seminvariants I, J, K, L of W can be expressed as

$$\begin{gathered} I = 2P_0\,, \qquad J = 4\,|\,P^i_j\,| = 2(P_0^2 - P_{00}), \\ K = 16\,|\,P^i_{j;1}\,| = 8(P_1^2 - P_{11}), \qquad L = 64\,|\,P^i_{j;2}\,| = 32(P_2^2 - P_{22}). \end{gathered} \tag{9.2}$$

To obtain expressions for the invariants Θ_w of W, we introduce the quantities Q^i_j obtained as follows.

From (6.3) we find by contraction

$$\bar{P}^i_i = \frac{1}{\Delta^2}(P^i_i + n\mu), \tag{9.3}$$

and eliminating μ from (9.3) and (6.3) gives us

$$\bar{Q}^i_j = \frac{1}{\Delta^2}Q^i_j, \qquad Q^i_j \equiv P^i_j - \frac{1}{n}\delta^i_j P^a_a. \tag{9.4}$$

Hence the Q^i_j are components of a relative tensor invariant. They satisfy the identity $Q^i_i = 0$.

If we define $Q_{\alpha\beta}$ by (7.1) with $P^i_{j;\alpha}$ replaced by $Q^i_{j;\alpha}$, then it is not difficult to show the following relations hold ($n = 2$):

$$\Theta_4 = 8Q_{00}\,, \qquad \Theta_6 = 32P_0Q_{00} - 40Q_{11} + 32Q_{02} = Q_{00}[36R_{11} + 2(X + 16P_0)],$$

$$\Theta_{10} = -64\begin{vmatrix} Q_{00} & Q_{01} \\ Q_{01} & Q_{11} \end{vmatrix} = 32Q_{00}^2R_{11}, \qquad \Theta_{18} = 2^{11}\begin{vmatrix} Q_{00} & Q_{01} & Q_{02} \\ Q_{01} & Q_{11} & Q_{12} \\ Q_{02} & Q_{12} & Q_{22} \end{vmatrix},$$

in which

$$X = 8\frac{d^2 \log Q}{dx^2} - \left(\frac{d \log Q}{dx}\right)^2, \qquad Q = |Q^i_j|, \qquad R^i_j = Q^{-1}Q^i_j, \qquad R_{11} = R^i_{j;1}R^j_{i;1}.$$

The invariant character of Θ_{10} follows immediately from the relations

$$\bar{R}^i_j = R^i_j, \qquad \bar{R}^i_{j;1} = \frac{1}{\Delta}R^i_{j;1} \tag{9.5}$$

under (6.1).

To obtain expressions for the semi-covariants of W we make use of the ϵ tensor with components[13] ϵ_{ij} which transform under (2.2) by $\bar{\epsilon}_{ij} = |a^i_j|^{-1}\epsilon_{kl}a^k_i a^l_j$.

The semi-covariants C, N, P, G take the form

$$C = 2\epsilon_{ij}P^i_k y^j y^k, \qquad N = 2\epsilon_{ij}Y^{jk}P^i_k, \qquad P = \epsilon_{ij}Y^i y^j, \qquad G = 2C_{;1},$$

where $y^{jk} = y^j y^k$, $Y^{jk} = y^{jk}_{;1}$. The quantity C being a relative semi-covariant, its covariant derivative is given by

$$C_{;1} = \frac{dC}{dx} + \tfrac{1}{2}L^i_i C.$$

We conclude with a generalization of another theorem in W, p. 116.

When in semi-canonical form, (2.1) can be written as

$$\frac{d^2y^i}{dx^2} + P^i_j y^j = 0, \tag{9.6}$$

and under (8.0) of the lemma with $x = \bar{x}$, this transforms to

$$\frac{d^2\bar{y}^i}{dx^2} + \bar{P}^i_j \bar{y}^j = 0. \tag{9.7}$$

It is possible to choose (8.0) so that $\bar{P}^i_i = 0$ in (9.7). For this purpose we take $h^i_j = \delta^i_j$ so that (8.3) becomes, on putting $\bar{M}^h_h = 0$, $M^i_k = P^i_k$, $\alpha = \Delta^{-\frac{1}{2}}$, and contracting on h and j,

$$n\alpha'' + \alpha P^i_i = 0. \tag{9.8}$$

This in turn can be reduced to

$$\mu = \frac{2}{n}P^i_i, \tag{9.9}$$

where μ is defined in (6.4); and a solution of (9.9) will lead to the desired transformation (8.0).

We thus have

[13] For the definition of these tensors see O. Veblen, *Invariants of Quadratic Differential Forms*, Chapter I.

THEOREM 9.1. *Every system* (2.1) *can be reduced to the form*

$$\frac{d^2 y^i}{dx^2} + Q^i_j y^j = 0,$$

where $Q^i_i = 0$.

Such a form is called a canonical form of (2.1).

NORTH CAROLINA STATE COLLEGE.

THE DISSECTION OF RECTANGLES INTO SQUARES

By R. L. Brooks, C. A. B. Smith, A. H. Stone and W. T. Tutte

Introduction. We consider the problem of dividing a rectangle into a finite number of non-overlapping squares, no two of which are equal. A dissection of a rectangle R into a finite number n of non-overlapping squares is called a *squaring* of R of *order* n; and the n squares are the *elements* of the dissection. The term "elements" is also used for the lengths of the sides of the elements. If there is more than one element and the elements are all unequal, the squaring is called *perfect*, and R is a *perfect rectangle*. (We use R to denote both a rectangle and a particular squaring of it.) Examples of perfect rectangles have been published in the literature.[1]

Our main results are:

Every squared rectangle has commensurable sides and elements.[2] (This is (2.14) below.)

Conversely, every rectangle with commensurable sides is perfectible in an infinity of essentially different ways. (This is (9.45) below.) (**Added in proof.** Another proof of this theorem has since been published by R. Sprague: Journal für Mathematik, vol. 182(1940), pp. 60–64; Mathematische Zeitschrift, vol. 46(1940), pp. 460–471.)

In particular, we give in §8.3 a perfect dissection of a square into 26 elements.[3]

There are no perfect rectangles of order less than 9, and exactly two of order 9.[4] (This is (5.23) below.)

The first theorem mentioned is due to Dehn, who remarked[5] that the difficulty of the problem is the semi-topological one of characterizing how the elements fit together. This is overcome here in §1 by associating a certain linear graph (the "normal polar net") with each "oriented" squared rectangle. The metrical properties of the squared rectangle are found to be determined by a certain flow of electric current through this network. Accordingly, in §2 we collect the relevant results from the theory of electrical networks. In particular, the elements of the squared rectangle can be calculated from determinants formed from the incidence matrix of the network. In §3, the elements are expressed in a different way, in terms of the subtrees of the network. This leads

Received May 7, 1940. We are indebted to Dr. B. McMillan, of Princeton University, for help with the diagrams.

[1] A bibliography is given at the end of this paper. Numbers in square brackets refer to this bibliography.

[2] Cf. [6], p. 319.

[3] This disproves a conjecture of Lusin; cf. [10], p. 272. For an independent example of a perfect square (published while this paper was in preparation) see [13].

[4] Partly confirming and partly disproving a conjecture of Toepken (see [18]).

[5] [12], p. 402.

to some relations between determinants and the subtrees of a network, and to some duality theorems. In §4, these duality theorems are applied to prove the converse of §1: that to any "polar net" corresponds a squared rectangle; and moreover, it is shown that (roughly speaking) the networks which correspond to the same squared rectangle in its two orientations are dual. In §5, the polar net is used to determine all the squared rectangles of a given order; in particular, the "simple" perfect rectangles of orders <12 are tabulated. §6 contains some theorems on the factorization properties of the elements of a squared rectangle, as determined in §2; as corollaries, we have some sufficient conditions for a squared rectangle to be perfect ((6.20), (6.21)). In §7, we give "non-uniqueness" constructions—in §7.1, of rectangles which can be dissected into the same elements in essentially different ways, and, in §7.2, of pairs of squared rectangles having the same shape but different elements. These constructions depend mostly on considerations of symmetry or duality in the corresponding networks. In §8, the results of §7.2 are used to give "perfect" squares; and in §9, a whole family of "totally different" perfect squares is worked out, and this leads to the result that every rectangle whose sides are commensurable is perfectible.

We conclude (§10) by outlining some generalizations—notably "rectangled rectangles", squared cylinders and tori, "triangulated" equilateral triangles, and "cubed cubes". We prove in particular that no "perfect" dissection of a rectangular parallelopiped into cubes is possible.[6]

1. The net associated with a squared rectangle

1.1. In any squaring of a rectangle R,[7] the sides of all the elements and of R will clearly be parallel to two perpendicular lines. We *orient* R by choosing one of these lines to be "horizontal" (i.e., parallel to the x-axis). The distinction between this configuration, and its reflections in the coördinate axes, is unimportant; but it is convenient to distinguish it from R in the other orientation (obtained by rotating R through an angle of $\frac{1}{2}\pi$), called the *conjugate* of R.

Consider the point-set formed by the horizontal sides of the elements of R. Its connected components will be horizontal line-segments (each consisting of a set of horizontal sides of elements of R); enumerate them as $p_1, \cdots, p_N$, say, where p_1, p_N are the upper and lower edges of R. Take N points $P_1, \cdots, P_N$ in the plane. Let E be an element of R; its upper edge will lie in some one of $p_1, \cdots, p_N$, say p_i: similarly, its lower edge will lie in p_j $(i \neq j)$. Join the points P_i, P_j by a line (simple arc) e. By taking all elements E of R, we get a network (linear graph) on $P_1, \cdots, P_N$ as vertices and the e's as 1-cells. Figure 1 provides an example.

The points P_1, P_N are the *poles* of the network. We can arrange the joins e in such a way that

[6] Answering a question raised by Chowla in [5].

[7] Throughout, all squares are supposed to have positive sides; thus zero elements are excluded.

(1.11) the network is realizable in a plane with no two 1-cells intersecting (except at a vertex).

(1.12) No circuit encloses a pole.

For we can realize the network as follows. Take P_i to be the mid-point of p_i; and take $\epsilon > 0$ sufficiently small. For each element E, take the vertical segment which bisects E, and cut off a length ϵ from each end, leaving a segment AB, say. Join the upper end of AB, A, to the P_i corresponding to the upper boundary of E, by a straight line-segment, and similarly join B to P_j corresponding to the lower edge of E. The path P_iABP_j is defined to be e. It is now easily verified that (1.11) and (1.12) hold.

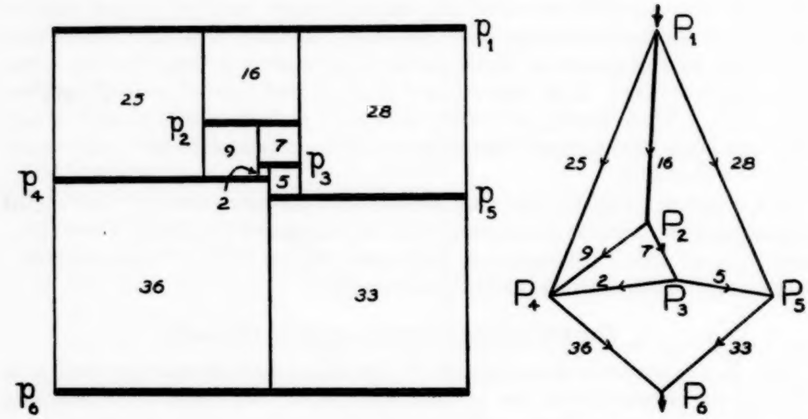

FIG. 1

Also we have clearly

(1.13) The network is connected.

Remark. In general there may be several 1-cells joining two vertices, though not if the squaring is perfect.

(1.14) DEFINITIONS. A network with more than one vertex, satisfying (1.11) and (1.13), is called a *net*. If two of the vertices of a net are assigned as "poles", and (1.12) is satisfied, the net is a *polar net* (p-net). The network constructed above is the *normal polar net* of the squared rectangle.

1.2. **Kirchhoff's laws.** With each 1-cell $e = P_iP_j$ of our normal p-net, associate the length of the side of the corresponding element E, directed from the "upper" point (P_i) to the "lower" point (P_j); call this the *current* in e. Then

(1.21) Except at the poles, the total current flowing into P_i is zero.

(For current flowing in = length of p_i = current flowing out.)

(1.22) The algebraic sum of the currents round any circuit is zero.

(For the current in a "wire" $e = P_iP_j$ is the vertical height of p_i above p_j.)

(1.23) The sum of the currents flowing into P_1 = length of horizontal side of R = sum of the currents flowing out of P_N.

(1.21) and (1.22) are the usual Kirchhoff laws for a flow of electric current in the net from P_1 to P_N, it being assumed that each 1-cell is a wire of unit conductance.

["Rectangulations" of rectangles can be dealt with similarly; the conductance of e will then be the ratio of the sides of E.]

Equations (1.21) and (1.22) can be interpreted differently. Consider the cellular 2-complex formed by embedding our p-net in a 2-sphere. We have on it a *Kirchhoff chain* (K-chain), viz., the 1-chain Σ (current in e)$\cdot e$. Then

(1.24) The K-chain is a cycle modulo its poles. (This $\leftrightarrow$ (1.21).)

(1.25) The K-chain is an absolute cocycle. (This $\leftrightarrow$ (1.22).)

2. Some results from the electrical theory of networks

2.1. In the previous section, we reduced the study of squared rectangles to the study of certain flows of electricity in networks. Here we collect the results on electrical networks in general which will be useful later.

Let $\mathfrak{N}$ be a connected network whose vertices are $P_1, \cdots, P_N$ ($N \geqq 2$). The 1-cells are called *wires*; there may be more than one wire joining two vertices, and there may be wires whose two ends coincide. With each wire is associated a positive real number, its *conductance*. We define a matrix $\{c_{rs}\}$ as follows:

(2.11) If $r \neq s$,

$$-c_{rs} = \begin{cases} \text{sum of conductances of all wires joining } P_r, P_s, \\ 0 \text{ if there are no such wires;} \end{cases}$$

$$c_{rr} = \text{sum of conductances of all wires joining } P_r \text{ to } other \text{ vertices.}$$

Thus

(2.12) $$c_{rs} = c_{sr}, \qquad \sum_r c_{rs} = 0.$$

We make the convention that if $\mathfrak{N}$ is explicitly called a *net*, all its conductances are 1. (The matrix $\{-c_{rs}\}$ is then the product of the usual incidence matrix of the oriented network, with its transpose.)

Let us return to the general case; from (2.12) we can readily show that all first cofactors of $\{c_{rs}\}$ are equal. We call their common value the *complexity* of the network, and denote it by C. It is known that $C > 0$. (An independent proof is given below; see (3.14).)

The second cofactor obtained by taking the cofactor of the component c_{su} in the cofactor of c_{rt} ($r \neq s$, $t \neq u$) is denoted by $[rs, tu]$. (If $N = 2$, $[12, 12] = 1 = -[21, 12]$.) We put $[rr, tu] = 0 = [rs, tt]$. The $[rs, tu]$'s are called the *transpedances* (generalized transfer impedances) of $\mathfrak{N}$.

Consider a flow of current from P_x to P_y (the *poles*). The currents in the wires satisfy (1.21); the potential differences (P.D.'s) satisfy the analogue of (1.22); and the total current I is given by (1.23). It is known[8] that these conditions (with Ohm's Law) determine the flow *uniquely* when I is given, and that

(2.13) P.D. from P_r to P_s when current I enters at P_x and leaves at P_y is $[xy,\, rs]\cdot I/C$.

It is convenient to take $I = C$, thus fixing the values of the currents and P.D.'s of the network. The flow with $I = C$ is called the *full flow*; and we speak of the "full currents", etc.

Applying this to the normal p-net of a squared rectangle, where all conductances are 1, so that all the transpedances are integers, we see from (1.21)–(1.23) and (2.13) that

(2.14) *Every squared rectangle has commensurable sides and elements.*

The H.C.F. of the full currents of a p-net is the *reduction* ρ of the p-net. Notice that ρ is also the H.C.F. of all the full P.D.'s of the p-net. The flow with $I = C/\rho$ is the *reduced* flow.

2.2. **Properties of the transpedances.** We have

$$[rs,\, tu] = [tu,\, rs] = -[sr,\, tu], \tag{2.21}$$

$$\sum_x c_{tx}\cdot[rs,\, tx] = C\cdot(\delta_{ts} - \delta_{tr}), \tag{2.22}$$

$$[rs,\, tu] + [rs,\, uv] = [rs,\, tv]. \tag{2.23}$$

(2.22) and (2.23) verify that (2.13) does in fact provide a solution of the Kirchhoff equations, and that the current at each pole is C.

We call $[rs,\, rs]$ the *impedance* of r, s, and write it $V(rs)$. Then

$$V(rs) = V(sr), \qquad V(rr) = 0, \tag{2.24}$$

$$2\cdot[rs,\, tu] = V(ru) + V(st) - V(su) - V(rt) \qquad \text{(from (2.23))}, \tag{2.25}$$

$$[rs,\, tu] + [tr,\, su] + [st,\, ru] = 0. \tag{2.26}$$

2.3. **Alterations to the network.** For later use, we need to know the effect on the transpedances of making certain alterations to the network $\mathfrak{N}$.

I. Introduce a new wire joining a vertex P_m of $\mathfrak{N}$ to a new vertex P_0. Let the new wire have conductance c; then, in the new network $\mathfrak{N}_1$,

$$C_1 = cC, \qquad V_1(m0) = C;$$

$$[ab,\, xy]_1 = c\cdot[ab,\, xy] \quad \text{if } 0 \neq a, b, x, y; \qquad [ab,\, m0]_1 = 0; \tag{2.31}$$

$$V_1(x0) = V_1(xm) + V_1(m0) = c\cdot V(xm) + C.$$

[8] [8], pp. 324–331.

These results are immediate from the definitions.

II. Identify two points P_x, P_y and ignore any wire that may have joined them. In the new network $\mathfrak{N}_2$,

$$C_2 = [xy, xy] = V(xy) \qquad \text{(from the definitions)}, \tag{2.32}$$

$$[rs, tu]_2 = \frac{[rs, tu]\cdot V(xy) - [rs, xy]\cdot[tu, xy]}{C} \tag{2.33}$$

(for these expressions satisfy Kirchhoff's laws for $\mathfrak{N}_2$, and agree with (2.32)). In particular,

$$V_2(rs) = \frac{V(rs)\cdot V(xy) - [rs, xy]^2}{C}. \tag{2.34}$$

((2.33) may be generalized as follows: C^n *divides the* $(n + 1)$*-th order determinants formed as minors of the matrix of transpedances.* This is an extension of the Cauchy-Sylvester identity.[9])

III. Introduce a new wire of conductance c in $\mathfrak{N}$, joining P_x and P_y. In the new network $\mathfrak{N}_3$ we have, from their definitions as determinants,

$$C_3 = C + c\cdot V(xy) = C + c\cdot C_2 \qquad \text{(from (2.32))}; \tag{2.35}$$

$$[rs, xy]_3 = [rs, xy]; \qquad \text{in particular, } V_3(xy) = V(xy). \tag{2.36}$$

Also

$$[rs, tu]_3 = [rs, tu] + c\cdot[rs, tu]_2\,; \tag{2.37}$$

for III is a combination of I and II. We introduce a new vertex P_0, join it to P_x by a wire of conductance c, and identify P_y and P_0. This enables us to verify (2.37).

3. Subtrees of a network: duality

We shall now characterize the complexity (and hence the transpedances) of a network more topologically, in terms of the "subtrees" of the network. This enables us to prove some duality theorems which will be useful later (§4) and are of interest in themselves.

3.1. As in the previous section, let $\mathfrak{N}$ be a connected network with conductances. By a *subnetwork* $\mathfrak{M}$ of $\mathfrak{N}$, we mean a network consisting of *all* the vertices of $\mathfrak{N}$ and some (or all) of the wires of $\mathfrak{N}$. A *subtree* of $\mathfrak{N}$ is a subnetwork which is a "tree"; i.e., is connected and has no circuits. Enumerate all the subtrees of $\mathfrak{N}$; let M_r be the product of the conductances of the wires of the r-th tree. Define H by:

$$H = \sum_r M_r\,. \tag{3.11}$$

[9] [19], p. 87.

When a new wire of conductance c is inserted joining P_x, P_y, let "H" for the new network be H_3; and when P_x, P_y are identified (as in §2.3, II), let "H" become H_2. Clearly,

$$H_3 = H + c \cdot H_2. \tag{3.12}$$

But this is the relation which holds between the complexities of these networks (2.35).

Also, for a connected network with only two vertices, C = sum of conductances of the wires joining P_1 to $P_2 = H$. Hence, by induction on the numbers of vertices and wires in $\mathfrak{N}$, we have:[10]

(3.13) THEOREM. *For any connected network with more than one vertex, having conductances assigned to the* 1-*cells,* $C = H$.

If the conductances are all positive, we clearly have $H > 0$. This proves

$$C > 0. \tag{3.14}$$

This interpretation of complexity in terms of trees enables us, if (2.32) is used, to express $V(xy)$ in terms of the trees of networks formed from $\mathfrak{N}$ by identifying certain pairs of its vertices, and hence in terms of the "tree-pairs" of $\mathfrak{N}$ (formed by omitting one wire from a subtree). Hence, using (2.25), we can get similar interpretations for all the transpedances.

In the case of a net, all conductances are 1, so H = number of subtrees of $\mathfrak{N}$; thus (3.13) gives an explicit formula for the number of subtrees of any connected network, in terms of the incidences of the network.

3.2. **Duality relations.** Now suppose that $\mathfrak{N}$ can be imbedded in a 2-sphere, and let $\mathfrak{N}^*$ be its dual on the sphere. The conductivity of a wire of $\mathfrak{N}^*$ is defined to be the reciprocal of that of the dual wire of $\mathfrak{N}$. Thus $\mathfrak{N}^{**} = \mathfrak{N}$, and the dual of a net is a net. The *codual* of a subnetwork $\mathfrak{M}$ of $\mathfrak{N}$ is the subnetwork $\mathfrak{M}^c$ of $\mathfrak{N}^*$ whose 1-cells are those *not* dual to any wire of $\mathfrak{M}$. Clearly $\mathfrak{M}^{cc} = \mathfrak{M}$.

It can be shown that

(3.21) A subnetwork $\mathfrak{M}$ of $\mathfrak{N}$ is a tree if and only if both $\mathfrak{M}$ and $\mathfrak{M}^c$ are connected.

Hence

(3.22) If $\mathfrak{M}$ is a subtree of $\mathfrak{N}$, then $\mathfrak{M}^c$ is a subtree of $\mathfrak{N}^*$; and conversely.

Let M_r^* equal the product of conductances of wires in the subtree (of $\mathfrak{N}^*$) which is codual to the r-th subtree of $\mathfrak{N}$. Let ω equal the product of conductances of all wires of $\mathfrak{N}$. Then, clearly,

$$M_r = \omega \cdot M_r^*. \tag{3.23}$$

[10] This result is due in principle to Kirchhoff ([9], p. 497). Cf. also [3].

Hence, using (3.22), (3.11), and (3.13), we have

(3.24) If C^* is the complexity of the dual of $\mathfrak{N}$, $\omega \cdot C^* = C$.

In particular, we have proved

(3.25) THEOREM. *Dual nets have equal complexities.*

3.3. **Polar duality.** Let $\mathcal{P}$ be a p-net. By (1.12), we can join the poles of $\mathcal{P}$ by an extra wire e_0, without violating (1.11). The resulting net $\mathcal{C}$ is called the *completed* net (c-net) of $\mathcal{P}$. Let $\mathcal{C}$ be imbedded in a 2-sphere, and let $\mathcal{C}^*$ be the dual of $\mathcal{C}$. From $\mathcal{C}^*$ omit e_0^*, the dual of e_0, and take the ends of e_0^* as poles. We get a p-net $\mathcal{P}'$, the *polar dual* of $\mathcal{P}$.[11]

Clearly $\mathcal{P}'' = \mathcal{P}$.

(The importance of polar duality arises from the fact that, as we shall show in §4.3, polar dual p-nets correspond to the same squared rectangle in its two "orientations" (§1.1).)

The p-dual (polar dual) of any 1-chain on $\mathcal{P}$ is defined in the obvious way (as having the same multiplicity on e_i^* as the given chain has on e_i).

(3.31) THEOREM. *The p-dual of the full Kirchhoff chain on a p-net $\mathcal{P}$ is the full Kirchhoff chain on the p-dual p-net $\mathcal{P}'$.*

Proof. We use $\mathcal{S}\mathfrak{N}$ to denote the cellular 2-complex formed by a network $\mathfrak{N}$ imbedded in a 2-sphere. F, δ are (as usual) boundary and coboundary operators, and * denotes duality with respect to the 2-sphere.

By (1.24), (1.25), the full K-chain $\mathcal{K}$ on $\mathcal{P}$ is a cycle relative to P_1, P_N (the poles of $\mathcal{P}$), and an absolute cocycle on $\mathcal{S}\mathcal{P}$. Hence, in $\mathcal{S}\mathcal{C}$ (where $\mathcal{C}$ is the completed net of $\mathcal{P}$) $\mathcal{K}$ is

(i) a relative cycle mod P_1, P_N, and

(ii) a relative cocycle mod the two 2-cells, say σ_1, σ_2, which have incidence with e_0, the "extra" join.

Dualizing, in $\mathcal{S}\mathcal{C}^*$, we see that $\mathcal{K}^*$ is

(i) a relative cocycle mod the 2-cells P_1^*, P_N^* and

(ii) a relative cycle mod σ_1^* and σ_2^*, the poles of $\mathcal{P}'$.

But $\mathcal{K}^*$ has zero multiplicity on e_0^*, for $\mathcal{K}$ has zero multiplicity on e_0. Hence $\mathcal{K}^*$ is (from (i)) a cycle on $\mathcal{P}'$ mod its poles, and (from (ii)) a cocycle on $\mathcal{S}\mathcal{P}'$ mod the 2-cell consisting of P_1^* and P_N^* together. But a single 2-cell cannot be a coboundary; for, dualizing, this would require a single vertex to be a boundary. Hence $\mathcal{K}^*$ is an absolute cocycle on $\mathcal{S}\mathcal{P}'$, besides being a cycle mod its poles. So $\mathcal{K}^*$ is a K-chain on $\mathcal{P}'$.

Let $\mathcal{K}'$ be the full K-chain on $\mathcal{P}'$; thus $\mathcal{K}^* = k \cdot \mathcal{K}'$, for some k.

[11] There may be several ways of placing e_0 on the sphere, and consequently several polar duals of $\mathcal{P}$ (differing, however, only trivially). We suppose that one of these is chosen arbitrarily. In the open plane, a convention will be introduced to make $\mathcal{P}'$ unique; cf. §§4.2, 4.3.

Let $\mathcal{P}$ have complexity C, and $V(1N) = V$. Let the corresponding numbers for $\mathcal{P}'$ be C', V'. Using (2.22) (with $c_{tz} = 1$), we have, in $\mathcal{P}$,

$$F(\mathcal{K}) = C\cdot(P_1 - P_N).$$

Therefore, in $\mathcal{SP}'$, $\delta(\mathcal{K}^*) = C\cdot(P_1^* - P_N^*)$ (these cells being oriented suitably). So

$$C = \begin{cases} \text{sum of currents around } F(P_1^*) \text{ in } \mathcal{K}^*, \\ \text{sum of currents along a path joining the end-points of } e_0^*, \\ \text{total P.D. between the poles of } \mathcal{P}', \text{ in the flow } \mathcal{K}^*. \end{cases}$$

Thus

$$C = k\cdot V'. \tag{3.32}$$

Similarly,

$$C' = (1/k)\cdot V. \tag{3.33}$$

Now, by (2.35), the complexity of $\mathcal{C}$ is $C + V$. Similarly, the complexity of $\mathcal{C}^*$ is $C' + V' = k\cdot(C + V)$, by (3.32), (3.33). But by (3.25) these complexities are equal. Hence $k = 1$, and $\mathcal{K}^*$ is the *full* K-chain on $\mathcal{P}'$.

4. The correspondence between p-nets and squared rectangles

4.1. We now sketch a proof showing that to each p-net corresponds a squared rectangle. This correspondence is many-one and is clarified by introducing the "normal form" of a p-net (§4.2). We can then set up a 1-1 correspondence between classes of p-nets (having the same normal form) and "oriented" squared rectangles, and can prove that p-dual p-nets correspond to "conjugate" squared rectangles. (Cf. §1.1.)

(4.10) LEMMA. *For a K-chain in a p-net $\mathcal{P}$, whose poles are P_1, P_N (suitably numbered),*

(4.11) *the potential of each vertex lies between the potentials of the poles;*

(4.12) *no currents go into P_1, or out of P_N;*

(4.13) *at a vertex P_i, there is an angle (in the plane) containing all ingoing currents, whose reflex contains all outgoing currents;*

(4.14) *on the boundary of a 2-cell of $\mathcal{SP}$, there are two vertices P_i, P_j such that no current round this boundary goes from P_j towards P_i.*

(We make the convention that zero currents do not go in or out.)

Proof. Let P_i be any vertex, and suppose a current goes into P_i. Then a current goes out of P_i along at least one wire, ending at P_j, say; and so on, until we reach a pole P_N (say). All this time the potential has been falling, so P_N is eventually reached; and the potential of P_i is thus not less than that of P_N. If all the currents at P_i are zero, we can connect P_i to a vertex P_k at which not all currents are zero, by a path of zero currents; and P_i, P_k have the same potential. Thus in all cases the potential of P_i is not less than that of P_N; and similarly it is not greater than that of P_1. This proves (4.11).

(4.12) follows at once from (4.11).

(4.13) has been proved for the poles; so let $i \neq 1, N$, and suppose that two outgoing currents at P_i separate (in the plane) two ingoing ones. As in the proof of (4.11), we can continue each of the first two wires into a path down to P_N, along which the current falls; and similarly we can extend the other two wires into paths of rising potential up to P_1. Hence one of the two former paths must intersect one of the latter again, say in P_j $(i \neq j)$. The potential of P_j is both less than and greater than the potential of P_i. This is a contradiction, and so (4.13) is proved.

(4.14) follows from (4.13) and (4.12) by dualizing, if we use (3.31).

4.2. **Normal form of a p-net.** Let $\mathcal{P}$ be a p-net imbedded in the open plane in such a way that its poles, P_1, P_N, can be joined in the "outside region" of $\mathfrak{S}\mathcal{P}$. (That is, $\mathcal{P}$ is first imbedded in the closed 2-sphere, an extra join e_0 of the poles is inserted, and the "point at infinity" is then taken to be in the 2-cell of $\mathfrak{S}\mathcal{P}$ which contains e_0.) We define the *normal form* of $\mathcal{P}$, as so placed in the plane, as follows:

Consider any (not identically zero) K-chain $\mathcal{K}$ on $\mathcal{P}$. Some currents may be zero; delete the corresponding wires, and delete all vertices at which all currents are zero. Since $C > 0$, we are left with a p-net still, having P_1, P_N as poles. Using (2.31), (2.37), (2.36) (with $c = 1$), we see that $\mathcal{K}$ is a K-chain for the new p-net $\mathfrak{N}$. Next take each *finite* 2-cell of $\mathfrak{S}\mathfrak{N}$, and consider the vertices on its boundary. By (4.14), the 2-cell with its boundary is homeomorphic to a convex polygon which has one highest point and one lowest point, and in which the potentials of the vertices increase with their heights. Moreover, they increase *strictly*; for now no currents are zero. Hence equipotential vertices on this boundary occur at most in pairs, which can all be respectively identified by a deformation across the 2-cell. Making all these identifications for all the finite 2-cells, we end with a p-net $\mathfrak{N}_0$, on P_1, P_N as poles, on which $\mathcal{K}$ is still a K-chain (by (2.33)). And there are now no two vertices at the same potential which can be joined without crossing some wire of $\mathfrak{N}_0$, or separating the poles in the "outside" region. In particular, there are no zero currents. $\mathfrak{N}_0$ is called the *normal form* of $\mathcal{P}$, in its given imbedding in the plane.

Notice that, while we have proved that $\mathcal{P}$, $\mathfrak{N}_0$ have the same *reduced* K-chains, they need not have the same full K-chains.

It is easily seen that the normal p-net of a squared rectangle is its own normal form.

4.3. We next prove

(4.31) THEOREM. *To every p-net $\mathcal{P}$ in the open plane corresponds a squared rectangle R, whose normal p-net is the normal form of $\mathcal{P}$. Polar dual p-nets correspond to conjugate squared rectangles.*

(The polar dual of a p-net $\mathcal{P}$ in the open plane is itself put in the open plane in the obvious way—e_0^* is taken to be in the "outside".)

Proof. Consider the full K-chain $\mathcal{K}$ on $\mathcal{P}$ and its dual, the full K-chain on

the p-dual net $\mathcal{P}'$. (By (3.31).) Let $\mathcal{P}$ have complexity C, and let the P.D. between its poles be V ($= V(xy)$). Thus ((3.32), (3.33)) the analogous numbers for $\mathcal{P}'$ are V and C respectively. We can take the lowest potentials in $\mathcal{P}$ and $\mathcal{P}'$ to be zero. Suppose a wire e in $\mathcal{P}$ has its end-points at potentials V_1, V_2, and its dual e^* has its end-points at potentials V_1', V_2'. If μ is a number such that $V_1 < \mu < V_2$, we say that *e comprises* $(\ , \mu)$; and if λ is such that $V_1' < \lambda < V_2'$, then *e comprises* $(\lambda, \ )$. If both relations are true, we say that *e comprises* (λ, μ).

Now, observing that $V_2 - V_1 =$ current in $e =$ current in $e^* = V_2' - V_1'$, we construct a squared rectangle R as follows: In a rectangle of height V and base C, we take, for each wire e of $\mathcal{P}$, the (closed) *square E* whose horizontal sides are at a height V_1, V_2 above the base (x-axis) and whose vertical sides are at a distance V_1', V_2' to the right of the left-hand vertical side (y-axis). If the current in e is zero, this square reduces to a single point, and is omitted.

Let $\lambda \neq$ any potential of a vertex of $\mathcal{P}'$, and $\mu \neq$ any potential of a vertex of $\mathcal{P}$. Then, if $0 < \lambda < C$, and $0 < \mu < V$, we have the following:

The wires (of $\mathcal{P}$) comprising $(\lambda, \ )$ form a single path from pole to pole, along which the direction of the current is constant. For, by (4.12) and duality, there is just one such wire terminating at each pole; and from (4.14), if one such wire carries current to a vertex, then just one such wire carries current from that vertex, and no more such wires terminate at that vertex.

Along this path, the potential increases steadily from pole to pole; also, by choice of λ, the currents along the path are non-zero. Hence just one wire in it comprises $(\ , \mu)$. So just one wire of $\mathcal{P}$ comprises (λ, μ). Thus the point of coördinates (λ, μ) belongs to just one of the squares E. It follows that the whole rectangle is filled completely and without overlap (except of boundaries of squares).

It is easy to see that the normal p-net of the squared rectangle so constructed is—to within reflection in the axes (which we always disregard)—the normal form of $\mathcal{P}$. Also, it is clear from the construction that the squared rectangle assigned to $\mathcal{P}$ differs from that assigned to $\mathcal{P}'$ only by interchange of horizontal and vertical; i.e., the two squared rectangles are conjugate.

In this way, we have a 1-1 correspondence between classes of p-nets in the plane having the same normal form, and "oriented" squared rectangles.

Definitions. As suggested by (4.31), the complexity of a p-net is called its (full) ***horizontal side*** (often written H instead of C); and the full P.D. between its poles is its ***vertical side*** (V). The "full elements" and "full sides" of a squared rectangle refer to those of its normal p-net. The "reduced elements" will be the same for all corresponding p-nets.

4.4. Defining a *cross* as a point of a squared rectangle which is common to four elements, and an "uncrossed" squared rectangle as one which has no crosses, we have:

The normal p-nets of uncrossed conjugate squared rectangles are p-duals.

For let $\mathcal{P}$ be the normal p-net of the squared rectangle R; and let $\mathcal{P}'$ be the p-dual of $\mathcal{P}$. Let $\mathcal{Q}$ be the normal p-net of the conjugate R' of R; thus, from

§4.3, $\mathcal{Q}$ is the normal form of $\mathcal{P}'$. Now, in deriving the normal form of $\mathcal{P}$ (as in §4.2) there are no zero currents to suppress; and there are no identifications of vertices possible, as otherwise R', and hence R, would have a cross. So $\mathcal{P}' = \mathcal{Q}$. That is, $\mathcal{P}$ and $\mathcal{Q}$ are p-duals.

(This result could be extended to crossed squared rectangles by making a suitable convention modifying the normal p-net when crosses are present; e.g., by regarding a cross as an "element of side zero".)

5. Enumeration of squared rectangles

5.1. **Computation.** To find all the squared rectangles of a given order n, we have only to make a list of all p-nets having n wires. There is no difficulty in this, if n is not too large. We can save some labor by noting that p-dual nets give essentially the same rectangles; also we can assume that no part of a net, not containing a pole, is joined to the rest only at one vertex. (For the currents in this part would all be zero, whereas we can restrict ourselves to "normal forms".) A convenient way of carrying out the calculations is to consider the c-nets. From each net of $n + 1$ wires, we remove one wire and take its end-points as poles in the remaining net (if it is a net; i.e., is connected). Dual c-nets give rise to pairs of polar dual p-nets; so we need consider only half the c-nets. The working can be simplified by a proper use of §2.2. In practice, the Kirchhoff equations are best solved directly (without using determinants); a single determinant then gives the *full* elements for all the p-nets derived from one c-net.

It follows from §2.3 that all p-nets derived as above from the same c-net will have the same (full) semiperimeter, viz., the horizontal side of the c-net; and that two p-nets which differ only in the choice of poles, and their (non-polar) duals, all have the same (full) horizontal sides, viz., the complexity of the nets. (By (3.25).) Thus a number which appears in the $(n + 1)$-th order as a side appears (several times) in the n-th order as a semiperimeter. These facts are illustrated in the table below (§5.3).

5.2. **The perfect rectangles of least order. "Simple" perfect rectangles**

(5.21) A squared rectangle which contains a smaller squared rectangle (and any p-net corresponding to it) is called *compound*; all other squared rectangles and p-nets are *simple*. A p-net $\mathcal{P}$, without zero currents, which has a part $\mathcal{Q}$ such that $\mathcal{Q}$ contains more than one wire, $\neq \mathcal{P}$, is joined to the rest at only two vertices Q_1, Q_2, and contains no pole (except perhaps for Q_1 or Q_2) is compound. For $\mathcal{Q}$ must be connected; and the squared rectangle corresponding to $\mathcal{P}$ will contain the smaller squared rectangle which corresponds to $\mathcal{Q}$ (with Q_1, Q_2 as poles).

(5.22) "*Trivial*" *imperfection.* If a p-net has two equal non-zero currents, it is *imperfect*, and these currents constitute an "imperfection". (This is equivalent to saying that the corresponding squared rectangle is not perfect.) If a p-net has a part, not containing a pole, joined to the rest by only two wires, or if it has a pair of vertices joined by two (or more) wires, these two wires will

clearly have equal currents. If these currents are non-zero, the resulting imperfection is said to be *trivial.* A p-net which has a non-trivial imperfection is called *non-trivially imperfect.* A non-trivially imperfect p-net may or may not have a trivial imperfection.

We now have the theorem:

(5.23) *The c-net derived from a simple perfect rectangle has no part (consisting of more than one wire and of less than all but one wire) joined to the rest at less than three vertices; and the same is true of its dual.*

For the normal p-net of the simple perfect squared rectangle (or of the conjugate squared rectangle) will otherwise have a zero current, or a trivial imperfection, or be compound.

A perfect rectangle of the smallest possible order must evidently be simple. Applying (5.23) to the method of §5.1, we readily find that

There are no perfect rectangles of order less than 9, *and exactly two perfect rectangles of order* 9.

Of the latter, one is well known;[12] the other is, we believe, new and has been drawn in Figure 1.

Below, we give a list of the simple perfect rectangles of orders 9–11. The compound perfect rectangles of these orders follows trivially.

5.3. Table of simple perfect rectangles.

Order	Full Sides	Semi-perimeter	Description of Polar Net (current from P_a to $P_b = ab$)	Reduction
9	66, 64	130	$ab = 30, ac = 36, bd = 14, cd = 8, be = 16, de = 2, ef = 18, df = 20, cf = 28.$	2
	69, 61	130	$ac = 25, ab = 16, ae = 28, bc = 9, bd = 7, dc = 2, de = 5, cf = 36, ef = 33.$	1
10	114, 110	224	$ab = 60, ac = 54, cb = 6, ce = 22, cd = 26, be = 16, ed = 4, bf = 50, ef = 34, df = 30.$	2
	130, 94	224	$ab = 44, ac = 38, ae = 48, cb = 6, ce = 10, cd = 22, ed = 12, bf = 50, df = 34, ef = 46.$	2
	104, 105	209	$ab = 60, ac = 44, cb = 16, cd = 28, bd = 12, be = 19, de = 7, bf = 45, ef = 26, df = 33.$	1
	111, 98	209	$ab = 44, ad = 26, ae = 41, dc = 11, de = 15, ce = 4, cb = 7, eb = 3, bf = 54, ef = 57.$	1
	115, 94	209	$ab = 34, ac = 19, ad = 23, ae = 39, cb = 15, cd = 4, de = 16, db = 11, bf = 60, ef = 55.$	1
	130, 79	209	$ab = 34, ac = 23, ad = 35, ae = 38, cb = 11, cd = 12, de = 3, bf = 45, df = 44, ef = 41.$	1

[12] First found, apparently, by Moroń [11]. See also [10], p. 272; [2], p. 93; [14], p. 8; and [4].

The full sides and semiperimeters of the simple perfect rectangles of the 11-th order are:

Order	Semi-perimeter	Sides
11	336	127, 209; 151, 185
	353	144, 209; 159, 194; 162, 191; 166, 187; 168, 185; 176, 177
	368	159, 209; 169, 199; 172, 196; 177, 191; 183, 185
	377	168, 209; 178, 199; 183, 194
	386	162, 224; 177, 209; 181, 205; 190, 196; 191, 195; 192, 194

Four of these are reducible, with reduction = 2; these are the rectangles whose sides are both even.

Of the 67 simple perfect rectangles of the 12-th order, eleven have reduction 2, eight have reduction 3, and one has reduction 4.

6. Theorems on reduction

In perfect rectangles of higher orders, much larger reductions occur; for example, a 19-th order rectangle with reduced sides 144 and 155 has $\rho = 80$. Its reduced elements are: $ab = 46$, $ad = 40$, $af = 28$, $ag = 41$, $bc = 10$, $bi = 36$, $ci = 26$, $dc = 16$, $de = 3$, $dh = 21$, $eh = 18$, $fe = 15$, $fg = 13$, $gk = 54$, $hl = 39$, $ij = 62$, $kj = 49$, $kl = 5$, $lj = 44$.

6.1. The following theorems on reduction are of interest.

(6.11) Theorem. *If one of the currents in a p-net is zero, the net is reducible.*

Let the poles be P_r, P_s, and the zero current be in a wire joining P_x, P_y. Then the transpedance $[rs, xy]$ is zero. On removing the wire in question (use (2.37) with $c = -1$, and (2.33)), the new value for $[rs, tu]$ is

$$[rs, tu]' = [rs, tu] - \frac{[rs, tu]\cdot V(xy) - [rs, xy]\cdot[tu, xy]}{C}$$

$$= [rs, tu]\cdot\frac{C - V(xy)}{C}.$$

Now, $C > C - V(xy) = C'$ (by (2.35)) > 0 (by (3.14)). Hence the H.C.F. of the $[rs, tu]$'s must be at least $C/(C - V(xy)) > 1$.

Definition. Let a positive integer $n = m\cdot k^2$, where m is square-free. Then k is called the *lower square root* of n, and mk is the *upper square root*.

(6.12) Theorem. *Let the full sides of a p-net be H, V. Then the reduction ρ is a multiple of the upper square root of the H.C.F. of H and V.*

By (2.34), remembering that $V_2(rs)$ is an integer, we have

$$C \text{ divides } V(rs)\cdot V(xy) - [rs, xy]^2.$$

Since $C = H$, and $V(rs) = V$ (taking P_r, P_s as poles), it follows that the H.C.F. of H, V divides $[rs, xy]^2$; whence the result.

(6.13) Corollary. *If the reduced sides of a squared rectangle have H.C.F. σ, then the reduction of any corresponding p-net is divisible by σ.*

(For, by (6.12), σ is a factor of the lower square root of the H.C.F. of the full sides and hence—since the lower square root divides the upper—of ρ.)

(An example is the rectangle 96×99 given in §7.1.)

(6.14) Corollary. *Any p-net of a squared square has for reduction a multiple of its reduced side.*

(6.15) Corollary. *A necessary and sufficient condition that a p-net be irreducible is that its two full sides be coprime.*

(6.16) Theorem. *All non-trivially imperfect p-nets are reducible.*

(6.17) Lemma. *If H, V, k are positive integers such that, for each positive integer n, $(H + nV, k) > 1$, then H, V, k all have a common factor greater than* 1.

Proof. Let N_0 be the product of all the primes which divide k but not H. (Empty product = 1.) Let p_0 be a prime factor of $(H + N_0V, k)$. Suppose $p_0 \nmid H$. Then $p_0 \mid N_0$. Hence, since $p_0 \mid (H + N_0V)$, we have $p_0 \mid H$, and this is a contradiction. So $p_0 \mid H$. Therefore p_0 divides N_0V but not N_0; so that p_0 divides V as well as H and k.

Proof of (6.16). Now let $\mathcal{P}$ be a p-net with full sides H $(= C)$ and V $(= V(1N))$; and let a non-trivial imperfection be $[1N, ab] = [1N, pq] = k$, say. Thus $k > 0$, and we do not have both $a = p$ and $b = q$. (Else the imperfection is trivial.)

Join P_1, P_N to produce the completed net $\mathcal{C}$. Let $\mathcal{Q}$ be the p-net formed from $\mathcal{C}$ by taking P_a, P_b as poles, and omitting one wire joining P_a, P_b. (Of course, there is such a wire; there may be several. It is easy to see, from considerations of "triviality", that $\mathcal{Q}$ is connected, and therefore a p-net.) Applying (2.33) to $\mathcal{C}$, and using (2.35), (2.36), (2.37), we have

$$(6.18) \qquad (H + V) \mid k \cdot (V(ab) - [ab, pq]),$$

where $V(ab)$, $[ab, pq]$ refer to the p-net formed by $\mathcal{C}$ with P_a, P_b as poles, and hence (2.36) refer equally well to $\mathcal{Q}$.

Now, we have $0 < V(ab) - [ab, pq] \leqq$ semiperimeter of $\mathcal{Q}$, with equality only if the current $[ab, qp]$ equals the total current of $\mathcal{Q}$. In this case, $\mathcal{C}$ must consist of two parts, joined only by the two wires P_aP_b and P_pP_q. Further, P_1, P_N, being joined in $\mathcal{C}$ by a wire not P_aP_b or P_pP_q, must lie in the same part. Hence the imperfection in $\mathcal{P}$ with which we started was trivial.

Hence (6.18) gives (since semiperimeter of $\mathcal{Q}$ = complexity of $\mathcal{C} = H + V$)

$$(6.19) \qquad (H + V, k) > 1.$$

Now let n be any positive integer. Join P_1, P_N by $n - 1$ extra wires (of unit conductance). The new p-net will have the same non-trivial imperfection (by (2.36)), so, applying (6.19) to the new net, and using (2.35), (2.32) repeatedly, we have

$$(H + nV, k) > 1.$$

The lemma (6.17) now shows that $(H, V) > 1$. Hence, by (6.15), $\mathcal{P}$ is reducible.

(6.20) COROLLARY. *All irreducible p-nets having no trivial imperfections give perfect squared rectangles.*

(6.21) COROLLARY. *If the complexity of a c-net is prime, all the squared rectangles derived from it (as in* §5.1) *will be perfect.*

These results are sometimes useful as tests for perfection.

For the reduced elements, we can prove (using the Euler polyhedron formula, and some consideration of the various cases)

(6.22) At least three of the reduced elements of any perfect rectangle are even. (Three is the best number possible.)

7. Construction of some special squared rectangles

7.1. **Conformal rectangles.** Two squared rectangles (or p-nets in this plane) which have the same shape (that is, have proportional sides) but are not merely rigid displacements of each other (in the case of p-nets, have not the same normal form) are called *conformal.* An example of a conformal pair is provided by the 9-th order rectangle 64×66 and a 12-th order rectangle of reduced sides 96, 99, whose (reduced) net is specified by: $ga = 31$, $ge = 21$, $gc = 44$, $ea = 10$, $ed = 11$, $ad = 1$, $dc = 12$, $ac = 13$, $ab = 27$, $cb = 14$, $bf = 41$, $cf = 15$.

Two conformal rectangles need not have the same full sides or reduction; for example, the rectangle 96×99 has reduction 3 (cf. (6.21)).

We now show how to construct conformal pairs having the *same* reduced elements (but differently arranged).

Suppose that a p-net $\mathcal{P}$ has a part $\mathcal{Q}$ joined to the rest only at vertices $A_1, \cdots, A_m$, say, and containing no pole different from an A_i. If $\mathcal{Q}$ has rotational symmetry about a vertex P, in which the A's are a set of corresponding points, then a simple symmetry argument shows that the potential of P (in $\mathcal{P}$) will be the mean of the potentials of $A_1, \cdots, A_m$. Hence if this is also true for another vertex P', P and P' will have equal potentials.

Coalesce P and P', forming (if this can be done in the plane) the p-net $\mathcal{P}_2$. If C is the complexity of $\mathcal{P}$, we see from (2.33) that, if $[ab, 1N]$ and $[ab, 1N]_2$ are corresponding elements in $\mathcal{P}$ and $\mathcal{P}_2$ (with 1, N referring to the poles),

$$[ab, 1N]_2 = \frac{V(PP')}{C}[ab, 1N].$$

Hence the elements of $\mathcal{P}_2$ are proportional to those of $\mathcal{P}$; $\mathcal{P}$ and $\mathcal{P}_2$ have the same reduced elements and sides. Their reductions are clearly in the ratio $C : V(PP')$. This construction enables conformal p-nets with the same elements to be written down.

A simple example is shown in Figure 2. Here A_1 and A_2 are poles. The rectangles are perfect and simple, and have reductions 5 and 6, and reduced sides 75 and 112.

In a more complicated example, illustrating a variation on the method, we make the potentials of three points P_1, P_2, P_3 equal. Although the network we start with is not planar, it becomes so when either P_1P_2 or P_1P_3 coincide. Such a network is specified below. It gives conformal simple perfect rectangles of the 28-th order, with reductions 96 and 120, reduced sides 6834 and 14065, and reduced elements: $A_1a = 3288$, $A_1P_1 = 3480$, $A_1b = 2512$, $A_1d = 2247$, $A_1i = 2538$, $aP_3 = 192$, $aA_3 = 3096$, $bP_3 = 968$, $bA_2 = 1544$, $P_1A_2 = 576$, $P_1A_3 = 2904$, $P_3c = 1160$, $A_2c = 584$, $cA_3 = 1744$, $de = 1014$, $dP_2 = 1233$, $eA_2 = 795$,

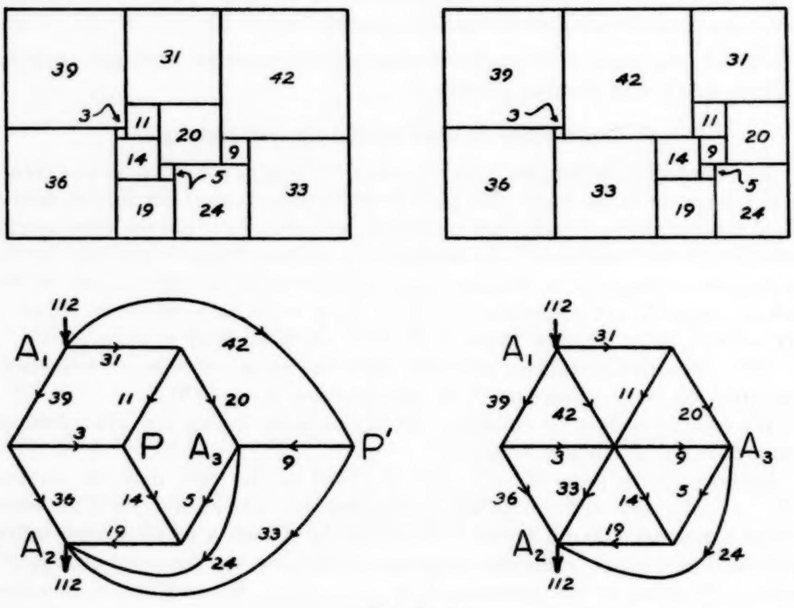

FIG. 2

$eP_2 = 219$, $iP_2 = 942$, $ih = 1596$, $P_2h = 654$, $P_2f = 579$, $P_2g = 1161$, $hA_3 = 2250$, $A_2f = 3$, $fg = 582$, $gA_3 = 1743$, $A_2A_3 = 2328$. (The poles are A_1 and A_3.)

These examples show that, even when the sides and elements of a simple perfect rectangle are given, the configuration is far from uniquely determined.

We now turn to the opposite problem of constructing conformal pairs of squared rectangles having *different* sets of elements. Again, symmetry considerations enable us to do this. We are led to pairs of rectangles (and p-nets) which are not merely conformal but have the same full sides. Such pairs are said to be *equivalent*.

7.2. **Symmetry method.** Let a p-net $\mathcal{P}$ have a part $\mathcal{Q}$ joined to the rest only at vertices $A_1, \cdots, A_m$, and containing no pole different from an A_i. Sup-

pose that $\mathcal{Q}$ has rotational symmetry in which the A's are a set of corresponding points, and that $\mathcal{Q}$ is not identical with its mirror-image. $\mathcal{Q}$ is the *rotor*, and the wires of $\mathcal{P} - \mathcal{Q}$ form the *stator*. In $\mathcal{P}$, replace $\mathcal{Q}$ by its mirror-image. It is easy to see that the full currents in the stator will be entirely unaffected, though (in general) the rotor currents will change. (This can be proved, e.g., by induction over the number of wires in the stator, if we use §2.) So we have, in general, a pair of equivalent rectangles, with different (though overlapping) sets of elements.

One of the simplest examples of this method is shown in Figure 3. This gives equivalent simple perfect rectangles of order 16, reduction 5, and reduced sides 671 and 504.[13]

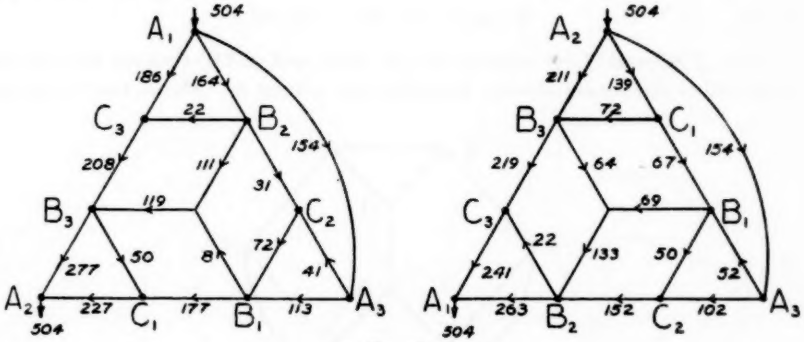

Fig. 3

We may generalize this method by noting that it remains effective when some of the A's are coincided (corresponding to the introduction of "wires of infinite conductance" in the stator). Or, again, we may take the stator to be itself a rotor, with $A_1, \cdots, A_m$ as its set of corresponding points (with possible coincidences). By reflecting both parts we can get pairs of equivalent rectangles having no elements in common.

7.3. **Special methods.** The preceding methods (and similar ones based on duality instead of symmetry) are useful for existence theorems, as in the next section; but other devices are more suitable for producing equivalent rectangles of small orders.

If, in a c-net $\mathcal{C}$, we can find two wires whose end-points—say P_a, P_b and P_x, P_y, respectively—satisfy

$$V(ab) = V(xy) \qquad \text{(in } \mathcal{C}\text{)}, \tag{7.31}$$

[13] The rotor of Figure 3 has a remarkable property. If currents I_1, I_2, I_3 (summing to zero) enter the *rotor* (considered as a net) at A_1, A_2, A_3, then the currents in B_3C_1, B_1C_2, B_2C_3 will be $I_1/7$, $I_2/7$, $I_3/7$, respectively. This explains the "extra" equalities of the currents in Figure 3. Other rotors of 15 wires (having the same type of symmetry) behave in a similar way. This phenomenon is not yet fully explained.

then the corresponding p-nets (obtained from $\mathcal{C}$ by omitting each of the two wires in turn, and taking its ends as poles) will be equivalent, if not identical. For they have the same semiperimeter in any case, viz., the complexity of $\mathcal{C}$.

By using the properties of symmetrical or self-dual networks, we can often demonstrate an equality like (7.31). For example, in Figure 4, it is clear that

$$V(gh) = V(cb) \tag{7.32}$$

and

$$[da, gh] = 0. \tag{7.33}$$

Hence (by (7.33) and (2.23) and symmetry)

$$[de, gh] = [ae, gh] = [de, cb]. \tag{7.34}$$

Now, (7.32) and (7.34) imply (if we use (2.37) and (2.33)) that the impedances of gh and cb remain equal when we add a wire joining de. Hence this new c-net

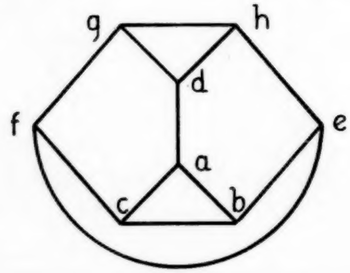

FIG. 4

satisfies (7.31), and so we get a pair of equivalent squared rectangles of the 12-th order. These rectangles are perfect, and provide the simplest example of equivalence among perfect rectangles. They both have reduction 2 and reduced sides 142 and 162. Their (reduced) specifications are respectively:

$gf = 57$, $gd = 85$, $dh = 77$, $de = 12$, $ad = 4$, $fe = 40$, $be = 13$, $eh = 65$, $ab = 3$, $ca = 7$, $cb = 10$, $fc = 17$; and $cf = 59$, $ca = 83$, $fe = 40$, $fg = 19$, $gh = 10$, $he = 11$, $gd = 9$, $dh = 1$, $ad = 4$, $de = 12$, $eb = 63$, $ab = 79$.

8. Construction of perfect squares

8.1. **Definition.** Two conformal rectangles are said to be *totally different* if C_2 times an element of the first is never equal to C_1 times an element of the second, where C_1, C_2 are their respective (corresponding) horizontal sides.

For equivalent rectangles this is equivalent to: No element of the first equals an element of the second.

A pair of totally different simple perfect squared rectangles gives us a perfect square at once; we have only to place them as in Figure 5, and add two corner

squares. This idea, though often in modified form, underlies all the constructions for perfect squares in this paper.

(8.11) It is easy to show (by the use of determinants) that if H, V and H', V' are the full sides of the rectangles used in this construction, then the resulting square will have full side $(H + V)\cdot(H' + V')$. In particular, if the rectangles are *equivalent*, the full side of the square is the square of an integer.

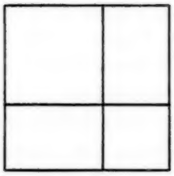

FIG. 5

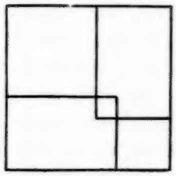

FIG. 6

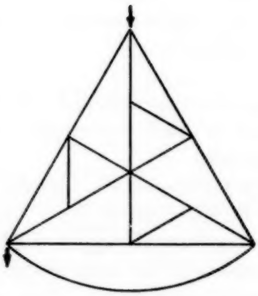

FIG. 7

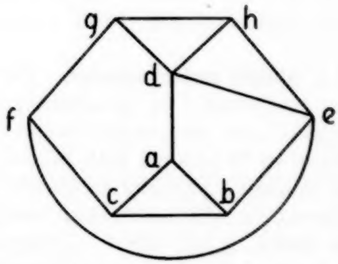

FIG. 8

8.2. **Symmetry method.** Equivalent perfect rectangles constructed as in §7.2 can be used to give us a perfect square. The stator is taken to be a single wire A_iA_j (drawn outside the rotor), one of whose end-points is a pole. The equivalent rectangles so obtained will have, *in general*,[14] just one element in common, the element corresponding to this stator. As this element is placed at a *corner* in both rectangles, we may "overlap" the rectangles as in Figure 6 to get a square.

One of the simplest perfect squares formed in this way is based on the rotor and stator shown in Figure 7. The square is of the 39-th order.

(8.21) It can be shown that, if H, V are the full sides of the equivalent rectangles used in this construction (§8.2), and E is the common element, then the full

[14] The "exceptional case", in which two elements from the following set: the rotor, its reflection, and the stator-element, are equal, seems in practice to be rare. It does occur, however, if the rotor has trivial imperfections, or if it has too much symmetry, or if it has triad symmetry and only 15 wires (cf. the previous footnote).

side of the resulting squared square is $(H + V - E)^2$, the square of an integer. In the case of *triad* symmetry ($m = 3$ in §7.2), we can show that $E \cdot (2H + 2V - E) = HV$, so that the full side of the squared square is, in this case, $H^2 + HV + V^2$.

8.3. **Perfect squares of smaller orders.** A perfect square of much smaller order is given by an elaboration of §7.3. We can show by an argument similar to that in §7.3, but longer, that in the net shown in Figure 8, $V(cf) = V(ge)$. (We use the facts that, if g and f are coalesced in Figure 8, the net becomes symmetrical and self-dual, and that Figure 8 results from Figure 4 by joining de.) Hence the two p-nets obtained by taking respectively c, f and g, e as poles in Figure 8 are equivalent (for their horizontal sides both equal the complexity). They are in fact perfect and totally different; and, though not both simple (the c, f one being obviously compound), the method of §8.1 is easily modified to give a perfect square, which is drawn in Figure 9. It is of the 26-th order. (The least possible order of a perfect square is unknown.)

We have also constructed, in a similar way, two perfect squares of the 28-th order, each of full side $(1015)^2$ and reduced side 1015.[15]

8.4. **Simple perfect squares.** The perfect squares constructed so far have all been compound. By generalizing the method of §8.2 to certain "squared polygons", we can obtain "simple" perfect squares.

First, let $\mathfrak{N}$ be a net with $A_1, \cdots, A_m$ as the vertices of its "outside" polygon, in order. Consider an electric flow in $\mathfrak{N}$ in which all of $A_1, \cdots, A_m$ are poles—i.e., in which currents I_i (not all zero) enter $\mathfrak{N}$ at A_i ($\sum I_i = 0$). Suppose that $I_i \geqq 0$ if $i > 1$. (This could be weakened; but some restriction on the order of the ingoing and outgoing currents is necessary.) Then the flow in $\mathfrak{N}$ corresponds to a squared polygon, of angles $\frac{1}{2}\pi$ and $\frac{3}{2}\pi$.

Proof. We reduce the number of poles of $\mathfrak{N}$ as follows: Suppose A_i is at potential V_i. Suppose there is more than one i for which $I_i > 0$; let $1'$, $2'$ be the least and second least such i's. If $V_{1'} = V_{2'}$, coalesce $A_{1'}$ and $A_{2'}$ (by joining them by a line outside the polygon $A_1 \cdots A_m$ and shrinking the line to a point); and let current $I_{1'} + I_{2'}$ enter there, the other currents being as before. The currents in $\mathfrak{N}$ will be unaltered, and there is now one fewer positive current entering the network. If $V_{1'} \neq V_{2'}$, we can suppose $V_{1'} > V_{2'}$. Join $A_{1'}$, $A_{2'}$ by a wire of conductance $I_{2'}/(V_{1'} - V_{2'})$ (passing outside the polygon $A_1 \cdots A_m$) and take currents $(I_{1'} + I_{2'})$ at $A_{1'}$, 0 at $A_{2'}$ and I_i at A_i for the other i's. Again, the currents in $\mathfrak{N}$ will be unaltered, and one fewer positive current enters the system. Repeating this process till there is only one positive external current left, we have the flow in $\mathfrak{N}$ "imbedded" in a flow with only two poles; in fact, in a p-net flow (except that some of the extra wires may have conductances different from 1). This corresponds to a "rectangled rectangle" R.

[15] See [16].

Stripping off the elements of R which correspond to the extra wires, we are left with a squared polygon, corresponding to $\mathfrak{N}$.

Since the currents I_i are (apart from sign) at our disposal, the shape of the squared polygon can be controlled. (It has $m - 2$ degrees of freedom.)

Now take for $\mathfrak{N}$ a pure rotor—i.e., a network having skew symmetry; and suppose that the points $A_1, \cdots, A_m$ are a set of corresponding points in $\mathfrak{N}$.

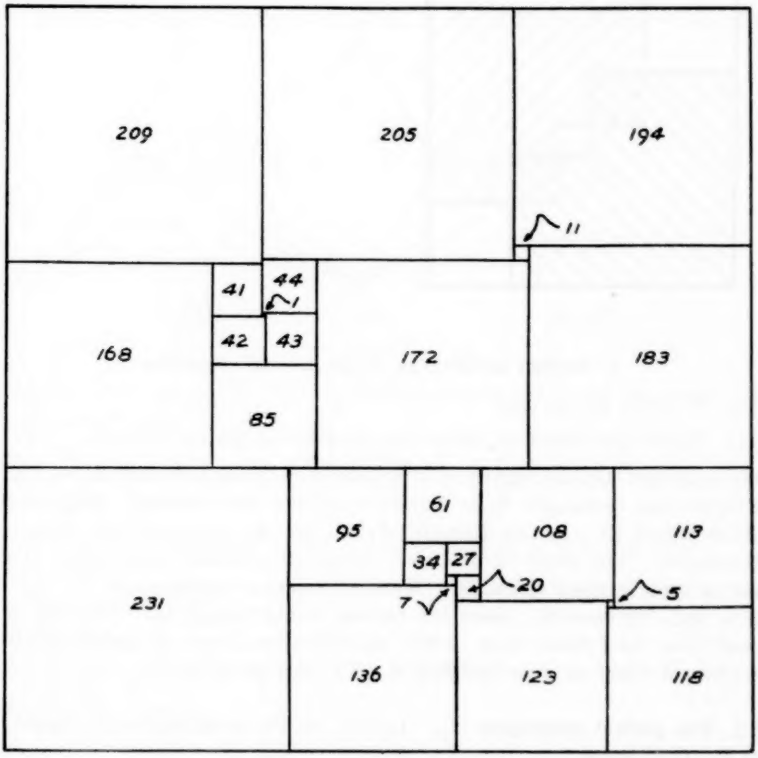

FIG. 9

If $\mathfrak{N}$ is replaced by its reflection (leaving the currents I_i invariant), the new squared polygon will have the same shape as the old—in fact, the two squared polygons will be "equivalent". For, as in §7.2, the rectangled rectangle R will be replaced by an equivalent one, in which the "extra" elements are the same as before.

By combining such a pair of equivalent polygons, as in Figure 10, and arranging their shape so that the overlapped portions coincide with elements

(which are then removed), and inserting three extra squares (in the center and at the corners), we can obtain a "simple" perfect square.

For instance, the rotor shown in Figure 11 gives rise to a simple "uncrossed" perfect square of order 55, which, when drawn out, disguises its symmetrical origin very skillfully.

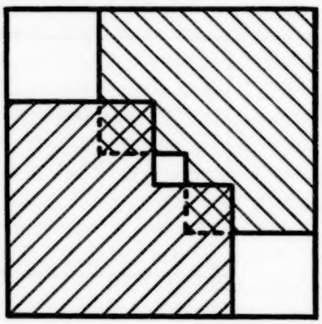

Fig. 10

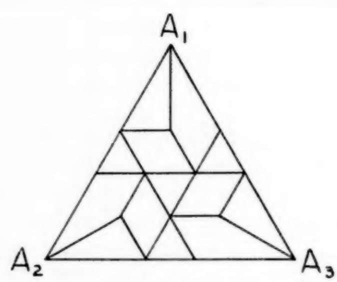

Fig. 11

9. Perfect subdivision of the general rectangle

9.1. We begin by proving:

(9.11) There exist infinitely many totally different perfect squares.

We construct such an aggregate of squares by the method of §8.2, taking for our equivalent rectangles those furnished by the "rotor-stator" diagram (cf. §7.2) of Figure 13. In this diagram, A_1, A_2 are the poles, and the wire A_1A_3 is the stator. The three "resistances" A_1B_2, etc., denote three copies of the p-net of some perfect rectangle. We shall select a sequence $\mathfrak{R}_n$ of suitable p-nets, and, for each $\mathfrak{R}_n$, form the corresponding square $\mathfrak{S}_n$. The sequence $\mathfrak{S}_n$ will then (as follows from (9.39)) have a subsequence of perfect squares, every two of which are totally different. This will prove (9.11).

9.2. **The perfect rectangles $\mathfrak{R}_n$.** Let $\mathfrak{R}_n$ be the p-net shown in Figure 12, with P_0, Q_0 as poles.

Write $\phi_r = [(2 + \sqrt{3})^r - (2 - \sqrt{3})^r]/2\sqrt{3}$. Thus

(9.21) ϕ_r is an integer; $\phi_0 = 0$; and $\phi_{r+1} - 4\phi_r + \phi_{r-1} = 0$.

It will readily be verified that a solution of Kirchhoff's equations is given by:

(9.22) Current in P_0P_r (from P_0 to P_r) is a_r, where

$$a_r = \tfrac{1}{2}\cdot[5\phi_n + \phi_{n-1} + 3\phi_r - 3\phi_{r-1}] \quad \text{if } 0 < r < n,$$

$$a_{n+1} = 3\phi_n .$$

Current in P_rQ_0 is b_r, where

$$b_r = \tfrac{1}{2}\cdot[5\phi_n + \phi_{n-1} - 3\phi_r - 3\phi_{r-1}] \quad \text{if } 0 < r < n + 1,$$

$$b_{n+1} = 2\phi_n + \phi_{n-1} .$$

Current in P_rP_{r+1} is c_r, where

$$c_r = 3\phi_r \qquad \text{if } 0 < r < n,$$

$$-c_n = \phi_n - \phi_{n-1}.$$

(This solution is in fact the full flow.)

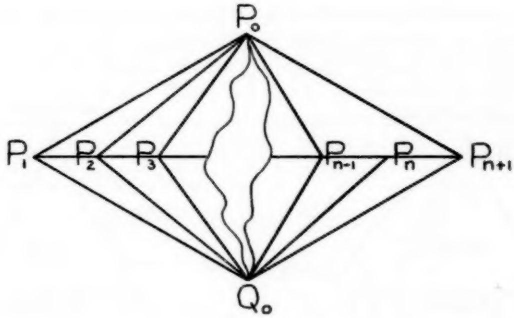

FIG. 12

Also the total current p_n (the horizontal side of $\mathfrak{R}_n$), and the total P.D. q_n (the vertical side) are given by:

(9.23) $$p_n = \tfrac{1}{2}\cdot[(5n + 1)\phi_n + (n + 2)\phi_{n-1}]; \qquad q_n = 5\phi_n + \phi_{n-1}.$$

Now, if $n > 2$, we see that

$$0 < c_1 < c_2 < \cdots < c_{n-2} < (-c_n) < c_{n-1} < b_n < b_{n+1} < b_{n-1} < b_{n-2}$$
$$< \cdots < b_1 < a_1 < a_2 < \cdots < a_{n-1} < a_{n+1}.$$

Hence

(9.24) If $n > 2$, $\mathfrak{R}_n$ is perfect.

From (9.23), we have

(9.25) q_n and $p_n/q_n \to \infty$ with n.

For later use, we note that

(9.26) $(p_n, q_n) \mid 9$.

Proof. From (9.23),

$$(n + 2)q_n - 2p_n = 9\phi_n \qquad \text{and} \qquad (5n + 1)q_n - 10p_n = 9\phi_{n-1}.$$

Now, we can prove by induction (using (9.21)) that $(\phi_n, \phi_{n-1}) = 1$. Thus (9.26) follows.

[(9.24) can be generalized: If in Figure 12 the wire P_0P_n is inserted and the wire P_0P_r removed, where $1 < r \leqq \tfrac{1}{2}n$, the resulting p-net $\mathfrak{R}_{nr}$ is perfect. $\mathfrak{R}_{n2}$

is essentially the same as $\mathfrak{R}_n$. The reduction ρ_r of $\mathfrak{R}_{nr}$ can be calculated; for instance, it can be shown that ρ_r is a factor of $(\phi_r - \phi_{r-1})$; and that $\rho_r = (\phi_r - \phi_{r-1})$ if and only if $n \equiv 0 \pmod{2r-1}$.]

9.3. We next prove

(9.31) THEOREM. *For all large n, the squared square $\mathfrak{S}_n$ is perfect.*

Consider the equivalent p-nets of Figure 13, where each "resistance" denotes a certain p-net $\mathfrak{R}$, of horizontal side p and vertical side q. (The other wires have conductance 1, as usual. Later, $\mathfrak{R}_n$ will be taken as $\mathfrak{R}$.)

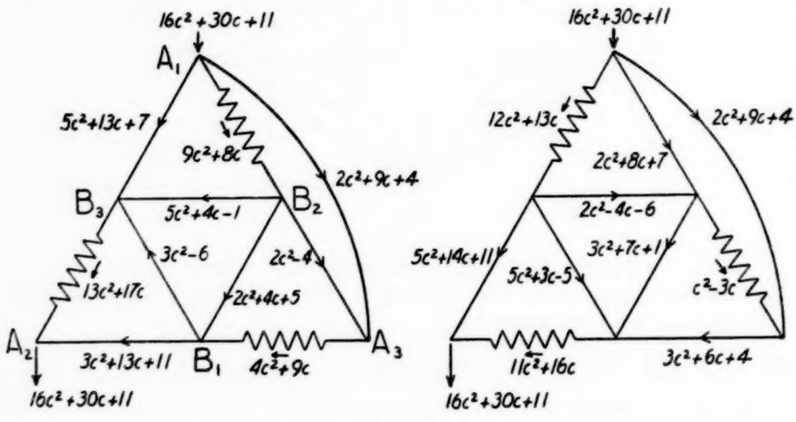

FIG. 13

Setting $c = p/q$ = effective conductance of these resistors, we find that the flows are as indicated in the diagram. (The quantities shown are currents.)

Hence, multiplying through by q^2, and adjoining the extra elements required in forming the squared square $\mathfrak{S}$ as in §8.2, we find that the elements of $\mathfrak{S}$ (some integral multiple of the reduced elements) are:

(9.32)
$$\left\{\begin{array}{ll} \text{(A)} & 14p^2 + 21pq + 7q^2,\ \ 5p^2 + 14pq + 11q^2,\ \ 5p^2 + 13pq + 7q^2, \\ & \qquad 5p^2 + 4pq - q^2,\ \ 5p^2 + 3pq - 5q^2, \\ & 3p^2 + 17pq + 20q^2,\ \ 3p^2 + 13pq + 11q^2,\ \ 3p^2 + 7pq + q^2, \\ & \qquad 3p^2 + 6pq + 4q^2,\ \ 3p^2 - 6q^2, \\ & 2p^2 + 9pq + 4q^2,\ \ 2p^2 + 8pq + 7q^2,\ \ 2p^2 + 4pq + 5q^2, \\ & \qquad 2p^2 - 4q^2,\ \ 2p^2 - 4pq - 6q^2. \\ \text{(B)} & \text{Multiples of the elements of } \mathfrak{R}\text{, the multipliers being respectively} \\ & 13p + 17q,\ 12p + 13q,\ 11p + 16q,\ 9p + 8q,\ 4p + 9q,\ p - 3q. \end{array}\right.$$

We also find that

(9.33) The side of $\mathfrak{S}$ is $19p^2 + 47pq + 31q^2$.

Now take $\mathfrak{R}$ to be $\mathfrak{R}_n$, so that $p = p_n$ and $q = q_n$; and let n be so large that (in virtue of (9.25))

(9.34) $$p_n > 180q_n .$$

We prove that, under this condition, $\mathfrak{S} = \mathfrak{S}_n$ is perfect.

(9.35) *The elements* (A) *are all different, and no element* (A) *equals an element* (B).

For the elements (A) in the above list are in strictly decreasing order; so no two of them are equal. Also the least element (A) is $2p^2 - 4pq - 6q^2$ which $> (13p + 17q)q$, which is greater than any element (B). Thus (9.35) follows.

(9.36) *No two elements* (B) *are equal.*

For suppose that two such elements are equal:

(9.37) $$\xi(\alpha p + \beta q) = \eta(\gamma p + \delta q),$$

where ξ, η are elements of $\mathfrak{R}_n$, and $\alpha p + \beta q$, $\gamma p + \delta q$ are two multipliers of (9.32). They are *different* multipliers; for $\mathfrak{R}_n$ is perfect by (9.24). Hence, by inspection of (9.32), $\alpha\delta - \beta\gamma \neq 0$. But, from (9.37), $(\alpha\xi - \gamma\eta)p = (\delta\eta - \beta\xi)q$. Hence

(9.38) $$p \mid (\delta\eta - \beta\xi)\cdot(p, q).$$

Now, if $\delta\eta - \beta\xi = 0$, we have (since $p \neq 0$) $\alpha\xi - \gamma\eta = 0$, and hence, if we eliminate ξ, η (which are not zero), it follows that $\alpha\delta - \beta\gamma = 0$. So we have $0 < |\delta\eta - \beta\xi| < 20q$ (by inspection of (9.32), since $0 < \xi, \eta < q$).

Hence, if we use (9.26), (9.38) gives $p < 180q$. This contradicts (9.34). And (9.31) now follows from (9.35) and (9.36); the squares $\mathfrak{S}_n$ are perfect, for large enough n.

(9.39) THEOREM. *Given any large enough n, then for all large enough N, $\mathfrak{S}_n$ and $\mathfrak{S}_N$ are totally different.*

Write $p_n = p$, $q_n = q$, $p_N = P$, $q_N = Q$. We bring $\mathfrak{S}_n$ and $\mathfrak{S}_N$ to the same size by multiplying the elements of $\mathfrak{S}_n$ (as given by (9.32)) by $19P^2 + 47PQ + 31Q^2$ and those of $\mathfrak{S}_N$ by $19p^2 + 47pq + 31q^2$. (This follows from (9.33).)

(9.40) *Each element* (B) *of $\mathfrak{S}_N$ is less than every element of $\mathfrak{S}_n$.*

For a typical element (B) of $\mathfrak{S}_N$ is

$$e = (\alpha P + \beta Q)\cdot(19p^2 + 47pq + 31q^2), \qquad \text{where } |\alpha|, |\beta| \leqq 17.$$

If n and N are large, this gives $e < 360Pp^2$. (This follows from (9.25).) But each element of $\mathfrak{S}_n$ is at least as large as P^2p (times some non-zero constant). Hence if $n >$ some n_0, and if then $N >$ some $N_0(n)$, so that P is large compared with p (see (9.25)), we have $e <$ each element of $\mathfrak{S}_n$.

(9.41) *Each element* (A) *of* $\mathfrak{S}_N$ *is greater than every element* (B) *of* $\mathfrak{S}_n$.

For any element (A) of $\mathfrak{S}_N$ is at least as large as P^2p^2 (times some non-zero constant), whereas an element (B) of $\mathfrak{S}_n$ is less than $360P^2p$.

(9.42) *No element* (A) *of* $\mathfrak{S}_N$ *can equal any element* (A) *of* $\mathfrak{S}_n$.

Otherwise we have

$$(aP^2 + bPQ + cQ^2)\cdot(19p^2 + 47pq + 31q^2) = (a'p^2 + b'pq + c'q^2)\cdot(19P^2 + 47PQ + 31Q^2),$$

where by (9.32) a, a', etc., are integers numerically less than 22. Hence

$$(9.43)\qquad P^2\cdot[(19a - 19a')p^2 + (47a - 19b')pq + (31a - 19c')q^2] = \text{similar terms in } PQ \text{ and } Q^2.$$

Now, $47a - 19b' \neq 0$; for otherwise $19 \mid a$, whereas $0 < a < 19$ (from (9.32)). Hence the left side of (9.43) is numerically at least as large as P^2pq (times some non-zero constant); in fact, if $a \neq a'$, it is as large as P^2p^2. But the right side of (9.43) is at most PQp^2 (times a constant). Hence, if N is taken large enough, so that P dominates both p and Q (this is possible, by (9.25)), (9.43) is impossible.

(9.40), (9.41), and (9.42) imply (9.39).

(9.44) COROLLARY. *There is a sequence* $\{\mathfrak{T}_n\}$ *of perfect squares, every two of which are totally different.*

This is immediate from (9.31) and (9.39) and proves (9.11).

A rough calculation shows that we may take $\mathfrak{T}_r = \mathfrak{S}_{10^{3(r+1)}}$. This could probably be greatly improved.

(9.45) THEOREM. *Any rectangle whose sides are commensurable can be squared perfectly in an infinity of totally different ways.*

Magnifying the rectangle suitably, we may suppose that its sides are integers h, k. Divide it into hk squares of side 1, by lines parallel to its sides. Take any positive integer n, and replace the i-th of these unit squares by $\mathfrak{T}_{nhk+i}$ (suitably contracted). By (9.44), this gives, for each n, a perfect subdivision of the given rectangle; and these subdivisions for any two values of n are "totally different".

Using the theorem of (2.14), we see that a rectangle can be squared *perfectly* if it can be squared at all.

It is plausible that any commensurable-sided rectangle can be squared perfectly and *simply*; possibly this can be proved in a similar way if we use some extension of §8.4; but this seems to involve laborious calculations.

10. Some generalizations

We mention briefly some of the extensions of the methods and results of this paper. A fuller discussion may perhaps appear later.

10.1. **Rectangled rectangles.** An immediate and natural generalization (as pointed out in §1.2) is to the problem of a rectangle dissected into a finite number of rectangles. The wires of the p-net merely have general (not necessarily equal) conductances.

There is also (cf. §8.4) a rather trivial extension in which the dissection is of a polygon (of angles $\frac{1}{2}\pi$ and $\frac{3}{2}\pi$). A more natural generalization, however, is given in the following section.

10.2. **Squared cylinders and tori.** We may regard a squared rectangle, after identification of its left and right sides, as a "trivial" example of a squared cylinder. The squared cylinders are found to correspond exactly to the relaxation of the condition (1.12) that no circuit of the p-net may enclose a pole. A second step brings us to the "squared torus". Using the existence theorem of (9.45), we can easily construct such figures. It is also possible to construct a *simple* non-trivial perfect torus; but this is not so easy.

Of course, the word "squared" may be replaced by "rectangled".

10.3. **Triangulations of a triangle.** In a rather different direction, we may consider dissections of a triangle into a finite number of triangles; particularly when all the triangles considered are *equilateral*. It is easily proved that *there is no perfect equilateral triangle*; i.e., that in any such dissection of an equilateral triangle into equilateral triangles, two of the latter are equal. Apart from this, the theory extends fairly completely. Duality relations, for example, are replaced by "triality" relations. We could also consider dissections into a mixture of equilateral triangles and regular hexagons, no two of these elements having equal sides; essentially this amounts to agglomerating the imperfections of an "equilateral triangled triangle" together by sixes. There is no difficulty in constructing such figures empirically, or in finding "perfect isosceles right-angled triangles"; however, it can be done by using the theory.

10.4. **Three dimensions.** We have seen that the "p-net" and its generalizations are satisfactory for plane dissections. As yet, however, there is no satisfactory analogue in three dimensions. The problem is less urgent, because *there is no perfect cube* (*or parallelopiped*). That is, in any dissection of a rectangular parallelopiped into a finite number of cubes ("elements"), two of the latter are equal.

Proof. It is easily seen that in any perfect rectangle, the smallest element is not on the boundary of the rectangle. Suppose we have a "perfect" cubed parallelopiped P. Let R_1 be its base. The elements of P which rest on R_1 "induce" a dissection of R_1 into a perfect rectangle. (We can clearly assume that more than one cube rests on R_1.) Let s_1 be the smallest element of R_1. Let c_1 be the corresponding element of P. Then c_1 is surrounded by *larger*, and therefore *higher*, cubes on all four sides; for, as remarked above, s_1 is surrounded by larger squares. Hence the upper face of c_1 is divided into a perfect

rectangle R_2 by the elements of P which rest on it; let s_2 be the smallest element of R_2; and so on. In this way, we get an infinite sequence of elements c_n of P, all different (for $c_{n+1} < c_n$). This is a contradiction.

This proof excludes generalizations of "perfect cylinders" to three (or more, a fortiori) dimensions; but it does not exclude the possibility of a *perfect three-dimensional torus* (product of three circles). It is not known whether such a thing can exist.

Bibliography

1. M. Abe, *On the problem to cover simply and without gap the inside of a square with a finite number of squares which are all different from one another*, Proceedings of the Physico-Mathematical Society of Japan, (3), vol. 14(1932), pp. 385-387.
2. W. W. Rouse Ball, *Mathematical Recreations*, 11th ed., New York, 1939.
3. C. W. Borchardt, *Ueber eine der Interpolation entsprechende Darstellung der Eliminations-Resultante*, Journal für Mathematik, vol. 57(1860), pp. 111-121.
4. S. Chowla, *Division of a rectangle into unequal squares*, Mathematics Student, vol. 7 (1939), p. 69.
5. S. Chowla, ibid., Question 1779.
6. M. Dehn, *Zerlegung von Rechtecke in Rechtecken*, Mathematische Annalen, vol. 57 (1903), pp. 314-332.
7. Jarenkiewycz, Zeitschrift für die mathematische und naturwissenschaftliche Unterricht, vol. 66(1935), p. 251, Aufgabe 1242; and solution, op. cit., vol. 68(1937), p. 43. (Also solutions by Mahrenholz and Sprague.)
8. J. H. Jeans, *The Mathematical Theory of Electricity and Magnetism*, Cambridge, 1908.
9. G. Kirchhoff, *Ueber die Auflösung der Gleichungen, auf welche man bei der Untersuchung der linearen Vertheilung galvanischer Ströme geführt wird*, Annalen d. Physik und Chemie, vol. 72(1847), p. 497.
10. M. Kraitchik, *La Mathêmatique des Jeux*, Brussels, 1930.
11. Z. Moroń, Przegląd Mat. Fiz., vol. 3(1925), pp. 152, 153.
12. A. Schoenflies-M. Dehn, *Einfuehrung in die analytische Geometrie der Ebene und des Raumes*, 2d ed., Berlin, 1931.
13. R. Sprague, Mathematische Zeitschrift, vol. 45 (1939), p. 607.
14. H. Steinhaus, *Mathematical Snapshots*, New York, 1938.
15. A. Stöhr, *Zerlegung von Rechtecken in inkongruente Quadrate*, Thesis, Berlin.
16. A. H. Stone, Question E. 401 and solution, American Mathematical Monthly, vol. 47 (1940).
17. H. Toepken, Aufgabe 242, Jahresberichte der deutschen Mathematiker-Vereinigung, vol. 47(1937), p. *2*.
18. H. Toepken, Aufgabe 271, op. cit., vol. 48(1938), p. *73*.
19. H. W. Turnbull, *Theory of Matrices, Determinants, and Invariants*, London, 1929.

Trinity College, Cambridge, England, and Princeton University.

THE LATTICE POINTS OF AN n-DIMENSIONAL TETRAHEDRON

By D. H. Lehmer

In this paper we consider the problem of determining the number of lattice points inside or on the boundary of the n-dimensional simplex or "tetrahedron" bounded by the n coördinate hyperplanes

$$x_1 = 0, x_2 = 0, \cdots , x_n = 0$$

and the hyperplane

$$\omega_1 x_1 + \omega_2 x_2 + \cdots + \omega_n x_n = \lambda,$$

where ω_i are positive and λ is a non-negative parameter. Points on the boundary are given the same weight as interior points. The total number of such points we denote by

$$N_n = N_n(\lambda) = N_n(\lambda \mid \omega_1 , \omega_2 , \cdots , \omega_n).$$

In other words, N_n is the number of sets $(x_1 , x_2 , \cdots , x_n)$ of non-negative integers for which the inequality

$$\omega_1 x_1 + \omega_2 x_2 + \cdots + \omega_n x_n \leqq \lambda \tag{1}$$

holds. Although the right triangle case ($n = 2$) has been considered by many writers,[1] there is as yet no published account of the general problem for $n > 2$. There are a number of isolated problems, however, which have been treated from time to time and which may be considered as special cases of the higher dimensional tetrahedron. It is the purpose of this paper to present a workable method for obtaining inequalities for the function $N_n(\lambda)$.

Three special tetrahedra may be mentioned as outstanding examples: (1) the equilateral tetrahedron, (2) the "additive" tetrahedron, in which the ω's are distinct integers, and (3) the "multiplicative" tetrahedron in which the ω's are logarithms of primes.

The first of these cases is the only one in which a really simple formula for $N_n(\lambda)$ can be given, and is useful for comparing approximate formulas. This case is interesting also as being that in which $N_n(\lambda)$ has the greatest discontinuity. The other two cases, which are interesting on account of their applications, will be considered briefly in what follows.

Before considering any special tetrahedra, however, we set down a fundamental recursion formula for the general tetrahedron obtained by dissecting the tetrahedron by the parallel hyperplanes

$$x_n = k \qquad (k = 0, 1, 2, \cdots).$$

Received June 1, 1940.

[1] See Koksma, *Diophantische Approximationen*, Berlin, 1936, pp. 102–110.

We obtain in this way a set of $(n - 1)$-dimensional tetrahedra, so that we may write

$$(2) \qquad N_n(\lambda \mid \omega_1, \cdots, \omega_n) = \sum_{k=0} N_{n-1}(\lambda - k\omega_n \mid \omega_1, \cdots, \omega_{n-1}),$$

or, what is the same thing,

$$(3) \quad N_n(\lambda \mid \omega_1, \cdots, \omega_n) - N_n(\lambda - \omega_n \mid \omega_1, \cdots, \omega_n) = N_{n-1}(\lambda \mid \omega_1, \cdots, \omega_{n-1}).$$

This exhibits $N_n(\lambda)$ as the solution of a linear difference equation in which $N_{n-1}(\lambda)$ is thought of as known.

In considering the equilateral tetrahedron there is no loss in generality in assuming that all the ω's are equal to unity.[2] Then by a simple induction on n it follows easily from (2) or (3) that

$$(4) \qquad N_n(\lambda \mid 1, 1, \cdots, 1) = \binom{[\lambda] + n}{n},$$

where $[\lambda]$ is the greatest integer not exceeding λ. In this case N_n is the number of "compositions" (rather than partitions) of the integers $\leqq \lambda$ as sums of not more than n positive integers. The step function $N_n(\lambda)$ has, for every n, discontinuities whenever λ is an integer m, and by (4) the jump is precisely

$$\binom{m + n - 1}{n - 1} = O(m^{n-1})$$

as $m \to \infty$. The ratio of this jump to the volume of the tetrahedron is asymptotic to n/m and hence tends to zero only rather slowly.

In the case of the additive tetrahedron, in which the ω's are distinct positive integers, $N_n(\lambda \mid \omega_1, \omega_2, \cdots, \omega_n)$ is clearly the total number of partitions of the numbers $\leqq \lambda$ into the parts $\omega_1, \omega_2, \cdots, \omega_n$. For many choices of the ω's it is possible to give exact formulas for $N_n(\lambda)$ as well as certain asymptotic results when λ is a function of n. This rather special topic will be dealt with in a separate note.

The most useful tetrahedron is the multiplicative one in which $\omega_k = \log p_k$, where $p_1, p_2, \cdots, p_n$ are distinct primes (especially the first n primes). Its importance is due to the following observation: Let P be a property of integers preserved under multiplication. Then if $p_1, p_2, \cdots, p_n$ have this property, at least

$$N_n(\log x \mid \log p_1, \log p_2, \cdots, \log p_n)$$

integers $\leqq x$ have P. The theory of numbers abounds with instances of properties P, and it is in many of these instances that it becomes desirable to find usable upper or lower bounds for $N_n(\lambda)$.

Quite recently in a Cambridge dissertation D. C. Spencer has considered the

[2] In the general tetrahedron one could, by a trivial change in the variable λ, assume that $\omega_1 = 1$, say.

problem of the n-dimensional tetrahedron. As approximating function he takes the polynomial

$$R_n(\lambda \mid \omega_1, \cdots, \omega_n) = \frac{B_n^{(n)}(\lambda + \Omega \mid \omega_1, \cdots, \omega_n)}{n!\, \omega_1 \omega_2 \cdots \omega_n},$$

where Ω denotes $\omega_1 + \omega_2 + \cdots + \omega_n$, and where $B_n^{(n)}(x \mid \omega_1, \cdots, \omega_n)$ is Nörlund's generalized Bernoulli polynomial,[3] and proves among other things that the difference

$$N_n(\lambda) - R_n(\lambda) \tag{5}$$

is $o(\lambda^{n-1})$ for arbitrary irrational ω's; that it is $o(\lambda^{n-1}/\log \lambda)$ for the multiplicative tetrahedron, $\omega_i = \log p_i$; and is $O((\log \lambda)^{n+\epsilon})$ for all $\epsilon > 0$ and for almost all $(\omega_1, \omega_2, \cdots, \omega_n)$.

The problem of getting actual numerical bounds for the difference (5) appears beset with grave practical difficulties, however. Even determining the sign of this difference in a particular case would seem to be nearly impossible.

The polynomial $R_n(\lambda \mid \omega_1, \cdots, \omega_n)$ is of the n-th degree in λ with coefficients which are complicated symmetric functions of the ω's. For $n = 5$, for example, we have

$$\begin{aligned} 5!\, \sigma_5 R_5(\lambda \mid \omega_1, \cdots, \omega_5) &= \lambda^5 + \tfrac{5}{2}\sigma_1\lambda^4 + \tfrac{5}{3}(\sigma_1^2 - \sigma_2)\lambda^3 + \tfrac{5}{2}\sigma_1\sigma_2\lambda^2 \\ &\quad + \{\tfrac{5}{6}(\sigma_2^2 + \sigma_1\sigma_3 - \sigma_4) + \tfrac{1}{6}s_4\}\lambda \\ &\quad - \tfrac{1}{12}\{\sigma_1^3\sigma_2 - 3\sigma_2^2\sigma_1 + 5\sigma_3\sigma_2 - \sigma_1^2\sigma_3 + \sigma_1\sigma_4 - 5\sigma_5\}, \end{aligned} \tag{6}$$

where σ_k denotes the sum of the products of the five ω's taken k at a time, and s_4 is the sum of their 4-th powers. This polynomial will be compared with others at the end of this paper.

Upper bounds for the general tetrahedron may be found in the form $Ae^{\alpha\lambda}$ by quite another method suggested by an inequality device used by Rankin.[4]

Let $\theta_i = e^{\omega_i}$ so that $\theta_i > 1$. Further let $\epsilon > 0$. Then the inequality (1) is equivalent to

$$\theta_1^{x_1} \theta_2^{x_2} \cdots \theta_n^{x_n} \leqq e^{\lambda}.$$

Hence we can write

$$\begin{aligned} N_n(\lambda \mid \omega_1, \cdots, \omega_n) &= \sum_{x_1\omega_1 + \cdots + x_n\omega_n \leqq \lambda} 1 = \sum_{\theta_1^{x_1} \cdots \theta_n^{x_n} \leqq e^{\lambda}} 1 \\ &\leqq e^{\epsilon\lambda} \sum (\theta_1^{x_1} \cdots \theta_n^{x_n})^{-\epsilon} < e^{\epsilon\lambda} \prod_{\nu=1}^{n} (1 - \theta_\nu^{-\epsilon})^{-1} = A_\epsilon e^{\epsilon\lambda}. \end{aligned} \tag{7}$$

[3] See N. E. Nörlund, *Differenzenrechnung*, Berlin, 1924, pp. 129–137.

[4] R. A. Rankin, *The difference between consecutive prime numbers*, London Mathematical Society Journal, vol. 13(1938), pp. 242–247.

As ϵ is made to approach zero, the product A_ϵ tends to infinity. The best results are obtained by making ϵ depend on λ in such a way that

$$\lambda = \sum_{\nu=1}^{n} \omega_\nu(\theta_\nu^\epsilon - 1)^{-1}.$$

For example, for the equilateral tetrahedron ($\omega_i = 1$) this condition reduces to $\epsilon = \log(1 + n/\lambda)$, and we find in this case that

$$N_n(\lambda \mid 1, 1, \cdots, 1) < (\lambda + n)^{\lambda+n}\, \lambda^{-\lambda} n^{-n} \sim (2\pi n)^{\frac{1}{2}} \binom{\lambda + n}{n} \left(1 + \frac{n}{\lambda}\right)^{-\frac{1}{2}}, \tag{8}$$

a result too large by a factor of nearly $(2\pi n)^{\frac{1}{2}}$.

The method presented herewith consists in constructing two approximating polynomials $P_n(\lambda \mid \omega_1, \omega_2, \cdots, \omega_n)$ and $Q_n(\lambda \mid \omega_1, \omega_2, \cdots, \omega_n)$ each of degree n in λ with coefficients depending on $\omega_1, \omega_2, \cdots, \omega_n$ and such that the inequalities

$$P_n(\lambda \mid \omega_1, \cdots, \omega_n) < N_n(\lambda \mid \omega_1, \cdots, \omega_n) < Q_n(\lambda \mid \omega_1, \cdots, \omega_n)$$

hold for all $\lambda \geqq 0$. There are, of course, infinitely many ways of doing this. The method adopted here is one in which $P_n(\lambda)$ and $Q_n(\lambda)$ are obtained recursively in such a way as to minimize, for all λ, the discrepancies between $N_k(\lambda)$ and its lower and upper bounds $P_k(\lambda)$ and $Q_k(\lambda)$. In other words, $P_1(\lambda)$, $P_2(\lambda)$, $\cdots$, $P_n(\lambda)$ are obtained as successive solutions of a sequence of linear difference equations, the n additive constants of these solutions being determined as the best possible. This is best accomplished by introducing Bernoulli polynomials.

We begin by stating a few facts, proved elsewhere,[5] about the maximum M_ν and the minimum m_ν of the ν-th Bernoulli polynomials $B_\nu(x)$ in the unit interval $0 \leqq x \leqq 1$. The notation employed is that in which

$$\begin{aligned} B_\nu(x) &= (B + x)^\nu = \sum_{k=0}^{\nu} \binom{\nu}{k} x^{\nu-k} B_k \\ &= x^\nu - \frac{\nu}{2} x^{\nu-1} + \frac{\nu(\nu-1)}{12} x^{\nu-2} - \frac{\nu(\nu-1)(\nu-2)(\nu-3)}{720} x^{\nu-4} + \cdots, \end{aligned}$$

where the notation for Bernoulli numbers B_k is such that symbolically

$$(B + 1)^\nu = B^\nu \qquad (B_0 = 1,\ B_1 = -\tfrac{1}{2},\ B_{2k+1} = 0,\ k > 0).$$

When ν is even, exact formulas for M_ν and m_ν may be given as follows:

$$\begin{aligned} M_0 &= m_0 = 1, \\ M_{4h} &= B_{4h}(\tfrac{1}{2}) = (1 - 2^{1-4h}) \mid B_{4h} \mid & (h > 0), \\ m_{4h} &= B_{4h}(0) = - \mid B_{4h} \mid & (h > 0), \\ M_{4h+2} &= B_{4h+2}(0) = B_{4h+2}\,, \\ m_{4h+2} &= B_{4h+2}(\tfrac{1}{2}) = -(1 - 2^{-1-4h}) B_{4h+2}\,. \end{aligned}$$

[5] American Mathematical Monthly, vol. 47(1940), pp. 533–538.

For ν odd, no exact formula for M_{2k+1} or m_{2k+1} exists. However, we have

$$M_{2k+1} = -m_{2k+1} < (2k+1)2^{-4k-2}\left\{|E_{2k}| + \frac{2}{\pi}(1-2^{-2k+1})|B_{2k}|\right\},$$

where E_{2k} are Euler numbers. Here the difference between the right side and M_{2k+1} tends rapidly to zero as $k \to \infty$. For $2k + 1 = 13$, for instance, this difference is less than 10^{-9}. A considerably simpler, though a little less precise, result is

$$M_\nu = -m_\nu < 2\nu!(2\pi)^{-\nu} \qquad (\nu \text{ odd}).$$

Here the right member when $\nu = 13$ exceeds the left by about $3\cdot 10^{-7}$. When $\nu < 13$, these results are less satisfactory and values of $M_{2k+1} = -m_{2k+1}$ as actually computed may be used. What we shall need, however, is not so much M_n and m_n as the functions

$$W_n = (B_n - M_n)/n \quad \text{and} \quad w_n = (B_n - m_n)/n.$$

These may be tabulated for $n \leqq 16$ as follows:

k odd		k even	
k	$w_k = -W_k$	k	w_k
		2	$.1250000000 = 1/2^3$
3	.01603750748	6	$.0078125000 = 1/2^7$
5	.004891638174	10	$.01513671875 = 31/2^{11}$
7	.003723587751	14	$.1666564941 = 5461/2^{15}$
9	.005283395737		
11	.012045150767		$-W_k$
13	.1611223890	4	$.0156250000 = 1/2^6$
15	.1856693827	8	$.00830078125 = 17/2^{11}$
		12	$.04217529297 = 691/2^{14}$
		16	$.8865060806 = 929569/2^{20}$

$$W_{4k+2} = w_{4k} = 0, \qquad W_1 = -1, \qquad w_1 = 0.$$

In what follows, we shall need estimates for sums of the type

$$S_k(x, \omega) = \sum_{1 \leqq \mu \leqq x/\omega} (x - \mu\omega)^k.$$

Hence we give the following lemma.

LEMMA. *If x and ω are positive, and if k is a non-negative integer, then*

$$(9) \qquad \omega^k(B_{k+1}(x/\omega) - M_{k+1}) \leqq (k+1)S_k(x, \omega) \leqq \omega^k(B_{k+1}(x/\omega) - m_{k+1}).$$

Proof. If in the well-known difference equation

$$(10) \qquad B_{k+1}(z+1) - B_{k+1}(z) = (k+1)z^k$$

we set $z = x/\omega - 1, x/\omega - 2, \cdots, x/\omega - [x/\omega]$ and add the resulting equations, we obtain

$$(k+1)\sum_{1\leqq\mu\leqq x/\omega}\left(\frac{x}{\omega}-\mu\right)^k = (k+1)S_k(x,\omega)\omega^{-k} = B_{k+1}\left(\frac{x}{\omega}\right) - B_{k+1}\left(\frac{x}{\omega}-\left[\frac{x}{\omega}\right]\right).$$

Since by definition

$$m_{k+1} \leqq B_{k+1}\left(\frac{x}{\omega}-\left[\frac{x}{\omega}\right]\right) \leqq M_{k+1},$$

the lemma now follows at once.

It is worth noting that for $k > 0$, either of the two equal signs in (9) holds for an infinity of x in arithmetic progression, in fact for

$$x = \omega(h + r_{k+1}) \qquad (h = 0, 1, 2, \cdots),$$

where r_{k+1} is that point on the interval $0 \leqq t < 1$ at which B_{k+1} attains the maximum or minimum value. For $k = 0$, however, the lemma states simply that

$$\frac{x}{\omega} - 1 \leqq \left[\frac{x}{\omega}\right] \leqq \frac{x}{\omega}. \tag{11}$$

Here the second equality sign holds infinitely often. The first never holds, but fails to do so by an arbitrarily small margin infinitely often.

We are now in a position to consider the problem of constructing $P_n(\lambda \mid \omega_1, \omega_2, \cdots, \omega_n)$. To begin with, we take the trivial case of $n = 1$ in which obviously

$$N_1(\lambda \mid \omega_1) = 1 + \left[\frac{\lambda}{\omega_1}\right].$$

By (11) the best possible choice of the polynomial $P_1(\lambda \mid \omega_1)$ is

$$P_1(\lambda \mid \omega_1) = \frac{\lambda}{\omega_1}. \tag{12}$$

Let us suppose that we have already constructed a polynomial of degree $k - 1$, say,

$$P_{k-1}(\lambda) = P_{k-1}(\lambda \mid \omega_1, \omega_2, \cdots, \omega_{k-1}) = \sum_{\nu=0}^{k-1} p_\nu^{(k-1)}\lambda^\nu$$

such that for all $\lambda > 0$

$$P_{k-1}(\lambda \mid \omega_1, \cdots, \omega_{k-1}) < N_{k-1}(\lambda \mid \omega_1, \cdots, \omega_{k-1}),$$

and where the coefficients $p_\nu^{(k-1)}$ depend only on $\omega_1, \omega_2, \cdots, \omega_{k-1}$. By (12),

$$p_0^{(1)} = 0, \qquad p_1^{(1)} = \frac{1}{\omega_1}. \tag{13}$$

To construct recursively the next polynomial $P_k(\lambda)$, we write the fundamental inequality (1) in the form

$$\omega_1 x_1 + \omega_2 x_2 + \cdots + \omega_{k-1} x_{k-1} \leqq \lambda - \omega_k x_k \qquad (k > 1).$$

Allowing x_k to range over the integers $0, 1, \cdots, [\lambda/\omega_k]$, we obtain by definition

$$N_k(\lambda \mid \omega_1, \omega_2, \cdots, \omega_k) = \sum_{0 \leqq \mu \leqq \lambda/\omega_k} N_{k-1}(\lambda - \mu\omega_k \mid \omega_1, \cdots, \omega_{k-1}).$$

Hence

$$\begin{aligned} N_k(\lambda) &> \sum_{0 \leqq \mu \leqq \lambda/\omega_k} P_{k-1}(\lambda - \mu\omega_k) = P_{k-1}(\lambda) + \sum_{\nu=0}^{k-1} p_\nu^{(k-1)} \sum_{1 \leqq \mu \leqq \lambda/\omega_k} (\lambda - \mu\omega_k)^\nu \\ &= P_{k-1}(\lambda) + \sum_{\nu=0}^{k-1} p_\nu^{(k-1)} S_\nu(\lambda, \omega_k). \end{aligned} \tag{14}$$

If we define for convenience the function $K_n(t)$ by

$$K_n(t) = \begin{cases} tm_n/n & \text{if } t \leqq 0, \\ tM_n/n & \text{if } t \geqq 0, \end{cases}$$

then the result of applying the lemma to (14) may be written

$$N_k(\lambda) > P_{k-1}(\lambda) + \sum_{\nu=0}^{k-1} p_\nu^{(k-1)} \frac{\omega_k^\nu}{\nu + 1} B_{\nu+1}(\lambda/\omega_k) - \sum_{\nu=0}^{k-1} \omega_k^\nu K_{\nu+1}(p_\nu^{(k-1)}). \tag{15}$$

The right member, which is a polynomial in λ of degree k, whose coefficients depend only on $\omega_1, \omega_2, \cdots, \omega_k$, we take for

$$P_k(\lambda) = P_k(\lambda \mid \omega_1, \cdots, \omega_k) = \sum_{\nu=0}^{k} p_\nu^{(k)} \lambda^\nu.$$

Expanding the Bernoulli polynomials and collecting the coefficients of the various powers of λ, we have the following recursion formula for $p_\nu^{(r)}$ or rather for

$$c_\nu^{(r)} = \omega_1 \omega_2 \cdots \omega_r p_\nu^{(r)}. \tag{16}$$

When $\nu > 0$

$$\nu c_\nu^{(k)} = c_{\nu-1}^{(k-1)} + \omega_k \frac{\nu}{2} c_\nu^{(k-1)} + \sum_{j=1}^{[\frac{1}{2}(k-\nu)]} \omega_k^{2j} A_j(\nu)\, c_{\nu+2j-1}^{(k-1)}, \tag{17}$$

where the coefficients

$$A_j(\nu) = \binom{\nu + 2j - 1}{2j} B_{2j} \tag{18}$$

do not depend upon k, and when once computed may be used in each successive determination of $P_k(\lambda)$ $(k = 1, 2, 3, \cdots, n)$. When $\nu = 0$, we have

$$c_0^{(k)} = \sum_{\nu=2}^{k-1} \omega_k^{\nu+1} V_{\nu+1}(c_\nu^{(k-1)}), \tag{19}$$

where

$$V_h(t) = \begin{cases} tW_h & \text{if } t \geqq 0, \\ tw_h & \text{if } t \leqq 0. \end{cases} \tag{20}$$

From (17) and (19) the $c_\nu^{(k)}$ may be found recursively starting from the initial values

$$c_1^{(1)} = 1, \qquad c_0^{(1)} = 0.$$

The first few polynomials $P_n(\lambda)$ are thus found to be

$$\begin{aligned}
\omega_1 P_1(\lambda \mid \omega_1) &= \lambda, \\
2!\omega_1\omega_2 P_2(\lambda \mid \omega_1, \omega_2) &= \lambda^2 + \omega_2\lambda, \\
3!\omega_1\omega_2\omega_3 P_3(\lambda \mid \omega_1, \omega_2, \omega_3) &= \lambda^3 + \tfrac{3}{2}(\omega_2 + \omega_3)\lambda^2 + \tfrac{1}{2}(\omega_3^2 + 3\omega_2\omega_3)\lambda + 3\omega_3^3 W_3, \\
4!\omega_1\omega_2\omega_3\omega_4 P_4(\lambda \mid \omega_1, \omega_2, \omega_3, \omega_4) &= \lambda^4 + 2(\omega_2 + \omega_3 + \omega_4)\lambda^3 \\
&\quad + (\omega_3^2 + \omega_4^2 + 3(\omega_2\omega_3 + \omega_3\omega_4 + \omega_2\omega_4))\lambda^2 \\
&\quad + (\omega_3^2\omega_4 + \omega_4^2\omega_2 + \omega_4^2\omega_3 + 3\omega_2\omega_3\omega_4 + 12\omega_3^3 W_3)\lambda \\
&\quad + 2\omega_4^3(3W_3(\omega_2 + \omega_3) + 2W_4\omega_4),
\end{aligned}$$

. . . .

$$\begin{aligned}
n!\omega_1\omega_2 \cdots \omega_n P_n(\lambda \mid \omega_1, \cdots, \omega_n) &= \lambda^n + \frac{n}{2}(\omega_2 + \omega_3 + \cdots + \omega_n)\lambda^{n-1} \\
&\quad + \frac{n(n-1)}{12}(\omega_3^2 + \cdots + \omega_n^2 + 3\sum \omega_2\omega_3)\lambda^{n-2} + \cdots.
\end{aligned} \tag{21}$$

It is seen that $P_n(\lambda \mid \omega_1, \cdots, \omega_n)$ is not a symmetric function of the ω's although $N_n(\lambda \mid \omega_1, \cdots, \omega_n)$ is. This raises the question as to the order in which the ω's should be introduced to maximize $P_n(\lambda \mid \omega_1, \cdots, \omega_n)$. That this choice of order will in general depend on λ is seen in the formula for $P_3(\lambda)$. In fact, if λ is large, the order should be obviously $\omega_1 \leqq \omega_2 \leqq \omega_3$ in spite of the fact that this will minimize the negative constant term $3\omega_3^3 W_3$. In general it is seen that for all large λ the value of $P_n(\lambda \mid \omega_1, \cdots, \omega_n)$ will be largest when

$$\omega_1 \leqq \omega_2 \leqq \cdots \leqq \omega_n$$

since this has the effect of maximizing the coefficients of λ^{n-1} and λ^{n-2}.

Another question that arises is: In what sense does $P_n(\lambda)$ approximate $N_n(\lambda)$? We may state the following answer.

THEOREM 1. *There exist for each $n > 0$ infinitely many n-dimensional tetrahedra for which*

$$N_n(\lambda) - P_n(\lambda) < C\lambda^{n-3},$$

where C is a positive constant depending on n.

Proof. Consider those n-dimensional equilateral tetrahedra in which $\omega_1 = \omega_2 = \cdots = \omega_n = 1$. By (4) we have

$$n!N_n(\lambda) = n!\binom{[\lambda] + n}{n} = \lambda^n + \left\{\frac{n(n+1)}{2} - n\delta\right\}\lambda^{n-1}$$
$$+ \left\{\frac{n(n-1)}{2}\,\delta^2 - \frac{n(n^2-1)}{2}\,\delta + \frac{n(n^2-1)(3n+2)}{24}\right\}\lambda^{n-2} + O(\lambda^{n-3}),$$

where we have written

$$[\lambda] = \lambda - \delta.$$

By (21), however, we have in this case

$$n!P_n(\lambda) = \lambda^n + \frac{n(n-1)}{2}\,\lambda^{n-1} + \frac{n(n-1)(n-2)(3n-1)}{24}\,\lambda^{n-2} + O(\lambda^{n-3}).$$

Hence

$$n!\{N_n(\lambda) - P_n(\lambda)\} = n(1-\delta)\lambda^{n-1} + \frac{n(n-1)}{2}\,(1-\delta)(n-\delta)\lambda^{n-2} + O(\lambda^{n-3}).$$

If we now choose λ of such a form that

$$1 - \delta = O(\lambda^{-2}),$$

as, for instance,

$$\lambda = k - k^{-2}, \qquad k \text{ an integer},$$

then

$$0 < N_n(\lambda) - P_n(\lambda) = O(\lambda^{n-3}).$$

This proves the theorem.

The polynomial $Q_n(\lambda)$ is constructed in a similar way. In fact since

$$N_1(\lambda \mid \omega_1) = 1 + \left[\frac{\lambda}{\omega_1}\right],$$

we may set

$$Q_1(\lambda \mid \omega_1) = 1 + \frac{\lambda}{\omega_1},$$

so that

$$Q_1(\lambda \mid \omega_1) \geqq N_1(\lambda \mid \omega_1),$$

the equality holding only when λ is an integer multiple of ω_1. If

$$Q_{k-1}(\lambda \mid \omega_1, \cdots, \omega_{k-1}) = \sum_{\nu=0}^{k-1} q_\nu^{(k-1)}\lambda^\nu$$

has already been determined in such a way that

$$Q_{k-1}(\lambda) \geqq N_{k-1}(\lambda \mid \omega_1, \cdots, \omega_{k-1}) \qquad (\lambda \geqq 0),$$

then

$$\begin{aligned} N_k(\lambda \mid \omega_1, \cdots, \omega_k) &= \sum_{\mu=0}^{[\lambda/\omega_k]} N_{k-1}(\lambda - \mu\omega_k \mid \omega_1, \cdots, \omega_{k-1}) \leqq \sum_{\mu=0}^{[\lambda/\omega_k]} Q_{k-1}(\lambda - \mu\omega_k) \\ &= Q_{k-1}(\lambda) + \sum_{\nu=0}^{k-1} \sum_{\mu=1}^{[\lambda/\omega_k]} q_\nu^{(k-1)} (\lambda - \mu\omega_k)^\nu \\ &= Q_{k-1}(\lambda) + \sum_{\nu=0}^{k-1} q_\nu^{(k-1)} S_\nu(\lambda, \omega_k). \end{aligned}$$

Applying our lemma we have

$$N_k(\lambda \mid \omega_1, \cdots, \omega_k) \leqq Q_{k-1}(\lambda) + \sum_{\nu=0}^{k-1} q_\nu^{(k-1)} \frac{\omega_k^\nu}{\nu+1} B_{\nu+1}(\lambda/\omega_k) + \sum_{\nu=0}^{k-1} \omega_k^\nu K_{\nu+1}(-q_\nu^{(k-1)}).$$

The polynomial on the right is taken as

$$Q_k(\lambda) = \sum_{\nu=0}^{k} q_\nu^{(k)} \lambda^\nu.$$

Expanding the Bernoulli polynomials as before and setting

$$d_\nu^{(k)} = \omega_1\omega_2 \cdots \omega_k q_\nu^{(k)},$$

we obtain recursion formulas for $d_\nu^{(k)}$ as follows:

For $\nu > 0$ we have corresponding to (17)

$$\nu d_\nu^{(k)} = d_{\nu-1}^{(k-1)} + \frac{\nu\omega_k}{2} d_\nu^{(k-1)} + \sum_{j=1}^{[\frac{1}{2}(k-\nu)]} \omega_k^{2j} A_j(\nu)\, d_{\nu+2j-1}^{(k-1)}, \tag{22}$$

where $A_j(\nu)$ is given in (18).

When $\nu = 0$ we have

$$d_0^{(k)} = \omega_k d_0^{(k-1)} - \sum_{\nu=0}^{k-1} \omega_k^{\nu+1} V_{\nu+1}(-d_\nu^{(k-1)}), \tag{23}$$

where the function $V_h(t)$ is given by (20).

Starting with $Q_1(\lambda \mid \omega_1)$ we find for the first few values of n

$$\begin{aligned} \omega_1 Q_1(\lambda \mid \omega_1) &= \lambda + \omega_1, \\ 2!\,\omega_1\omega_2 Q_2(\lambda \mid \omega_1, \omega_2) &= \lambda^2 + (2\omega_1 + \omega_2)\lambda + \tfrac{1}{4}\omega_2^2 + 2\omega_1\omega_2, \\ 3!\,\omega_1\omega_2\omega_3 Q_3(\lambda \mid \omega_1, \omega_2, \omega_3) &= \lambda^3 + \tfrac{3}{2}(2\omega_1 + \omega_2 + \omega_3)\lambda^2 \\ &\quad + (6\omega_1\omega_2 + 3\omega_1\omega_3 + \tfrac{3}{2}\omega_2\omega_3 + \tfrac{3}{4}\omega_2^2 + \tfrac{1}{2}\omega_3^2)\lambda \\ &\quad + 6\omega_1\omega_2\omega_3 + \tfrac{3}{4}\omega_2^2\omega_3 + \tfrac{3}{4}\omega_3^2\omega_1 + \tfrac{3}{8}\omega_3^2\omega_2 + \frac{3^{\frac{1}{2}}}{36}\omega_3, \end{aligned}$$

. . . .

$$
\begin{aligned}
n!\omega_1 \cdots \omega_n Q_n(\lambda \mid \omega_1, \cdots, \omega_n) &= \lambda^n + \frac{n}{2}(2\omega_1 + \omega_2 + \cdots + \omega_n)\lambda^{n-1} \\
&\quad + \frac{n(n-1)}{2}\{2\omega_1\omega_2 + \omega_1(\omega_3 + \cdots + \omega_n) + \tfrac{1}{2}\sum \omega_2\omega_3 + \tfrac{1}{4}\omega_2^2 \\
&\quad + \tfrac{1}{6}(\omega_3^2 + \cdots + \omega_n^2)\}\lambda^{n-2} + \cdots .
\end{aligned} \tag{24}
$$

As in the case of $P_n(\lambda)$, it is seen that if $\omega_1 \leqq \omega_2 \leqq \cdots \leqq \omega_n$, the best results are obtained for large λ since this minimizes the coefficients of λ^{n-1} and λ^{n-2}.

As an analogue of Theorem 1 we have the less satisfactory result.

THEOREM 2. *There exist for each $n > 0$ infinitely many n-dimensional tetrahedra for which*

$$Q_n(\lambda) - N_n(\lambda) = \frac{\lambda^{n-2}}{8(n-2)!} + O(\lambda^{n-3}).$$

Proof. As before take the equilateral case

$$\omega_1 = \omega_2 = \cdots = \omega_n = 1$$

and let λ be an integer so that

$$n!N_n(\lambda) = \lambda^n + \frac{n(n+1)}{2}\lambda^{n-1} + \frac{n(n^2-1)(3n+2)}{24}\lambda^{n-2} + O(\lambda^{n-3}),$$

whereas by (24)

$$n!Q_n(\lambda) = \lambda^n + \frac{n(n+1)}{2}\lambda^{n-1} + \frac{n(n-1)}{24}(3n^2+5n+5)\lambda^{n-2} + O(\lambda^{n-3}).$$

The theorem now follows at once from subtracting the right sides.

The fact that the coefficients of $P_n(\lambda)$ and $Q_n(\lambda)$ are complicated functions of $\omega_1, \omega_2, \cdots, \omega_n$ does not mean that the actual values of these coefficients cannot be found readily when numerical values of $\omega_1, \omega_2, \cdots, \omega_n$ are given. In fact the recurrence formulas (17), (19), (22) and (23) enable one to compute readily the successive numerical values of the coefficients $c_\nu^{(k)}$ and $d_\nu^{(k)}$, and hence $p_\nu^{(k)}$ and $q_\nu^{(k)}$. It has been quite feasible to compute these coefficients up to as high as $n = 13$ and 14, in connection with an investigation into the first case of Fermat's last theorem,[6] which involved the multiplicative tetrahedron.

A quite valuable check at each stage of the work is afforded by

THEOREM 3.

$$P_k(-\omega_k) = P_k(0) - P_{k-1}(0), \tag{25}$$

$$Q_k(-\omega_k) = Q_k(0) - Q_{k-1}(0). \tag{26}$$

[6] To appear shortly in the Bulletin of the American Mathematical Society.

Proof. To prove the first relation substitute first $\lambda = -\omega_k$ and then $\lambda = 0$ into the right member of (15), and subtract the results obtained so as to get

$$
\begin{aligned}
P_k(-\omega_k) - P_k(0) &= P_{k-1}(-\omega_k) - P_{k-1}(0) \\
&\quad + \sum_{\nu=0}^{k-1} p_\nu^{(k-1)} \frac{\omega_k^\nu}{\nu+1} \{B_{\nu+1}(-1) - B_{\nu+1}(0)\}.
\end{aligned} \tag{27}
$$

But from (10) with $k = \nu$ it follows that

$$B_{\nu+1}(-1) - B_{\nu+1}(0) = -(-1)^\nu(\nu + 1).$$

Hence the sum in the right member of (27) becomes

$$-\sum_{\nu=0}^{k-1} p_\nu^{(k-1)}(-\omega_k)^\nu = -P_{k-1}(-\omega_k),$$

so that (25) follows at once. (26) follows in precisely the same way.

In considering the multiplicative case J. B. Rosser[7] obtained a lower bound for $N_n(\lambda)$ as a polynomial $f_n(\lambda)$ which, when extended to the general tetrahedron, may be written

$$f_n(\lambda) = \frac{1}{n!\,\omega_1\omega_2 \cdots \omega_n}\left(\lambda^n + \frac{n}{2}\sigma_1'\lambda^{n-1} + \frac{n(n-1)}{2^2}\sigma_2'\lambda^{n-2} + \cdots + \frac{n!}{2^{n-1}}\sigma_{n-1}'\lambda\right),$$

where σ_k' is the sum of the products k at a time of $\omega_2, \omega_3, \cdots, \omega_n$. The first two coefficients of $f_n(\lambda)$ will be seen to agree with those of $P_n(\lambda)$. In fact

$$P_n(\lambda) - f_n(\lambda) = \frac{\lambda^{n-2}}{12(n-2)!\,\omega_1\omega_2 \cdots \omega_n}(\omega_3^2 + \omega_4^2 + \cdots + \omega_n^2) + O(\lambda^{n-3}).$$

Hence for $n > 2$ and $\lambda > \lambda_0$

$$f_n(\lambda) < P_n(\lambda) < N_n(\lambda).$$

By way of comparison of the various approximations to $N_n(\lambda)$ discussed above, we give the actual polynomials in the typical example of $N_5(\lambda \mid \log_{10} 2, \log_{10} 3, \log_{10} 5, \log_{10} 7, \log_{10} 11)$, i.e., the number of positive integers $\leq 10^\lambda$ divisible by no prime exceeding 11, and compare their values at several points with the exact values of $N_5(\lambda)$, as kindly furnished by Dr. A. E. Western, who has prepared extensive tables of $N_n(\lambda)$ in the multiplicative case. From these tables he has constructed an approximating polynomial $\phi_n(\lambda)$ by applying the method of least squares. The polynomial ϕ_5 is also given below, and compared with the others.

[7] *On the first case of Fermat's last theorem*, Bulletin of the American Mathematical Society, vol. 45(1939), pp. 636–640.

$$R_5(\lambda) = .094319\lambda^5 + .79313\lambda^4 + 2.46300\lambda^3 + 3.59621\lambda^2 - 6.36020\lambda - .037937,$$

$$P_5(\lambda) = .094319\lambda^5 + .72215\lambda^4 + 1.97819\lambda^3 + 2.26936\lambda^2 + .87536\lambda - .082148,$$

$$Q_5(\lambda) = .094319\lambda^5 + .86411\lambda^4 + 3.03689\lambda^3 + 5.27786\lambda^2 + 5.01395\lambda + 2.69600,$$

$$f_5(\lambda) = .094319\lambda^5 + .72215\lambda^4 + 1.61864\lambda^3 + 1.17723\lambda^2 + .20762\lambda,$$

$$\phi_5(\lambda) = .0033629\lambda^5 + 5.14087\lambda^4 - 71.79074\lambda^3 + 596.2170\lambda^2 - 2245.997\lambda + 3327.38.$$

The following table gives the exact values of $N_n(\lambda)$ together with the discrepancies of the approximating polynomials.

λ	$N_5(\lambda)$	$N - R$	$N - P$	$Q - N$	$N - f$	$\phi - N$
1	10	9.45	4.13	6.98	6.18	589.95
2	55	17.96	13.86	19.97	22.36	651.29
3	192	25.09	34.21	48.15	55.67	209.22
5	1197	40.60	142.61	177.14	218.11	.53
8	7838	58.38	624.42	727.55	884.53	1.65
10	20193	70.77	1325.78	1497.59	1801.20	140.41
10.5	24932	72.75	1567.04	1761.97	2110.69	239.94

To illustrate the use of equation (7) in this case suppose we attempt to represent $N_n(\lambda)$ by an exponential function near the value 20193. Then the λ of equation (7) is $10 \log_e 10 = 23.026 \cdots$. The best value of ϵ is found to be about .18889, and the value of A is about 14797. Hence

$$N_5(\lambda) < 14797 \cdot 10^{.18889\lambda}.$$

Substituting $\lambda = 10$ and 10.5 we find 114540 and 142360, values which are too large by the factors 5.67 and 5.71 respectively, which do not differ greatly from $(2\pi n)^{\frac{1}{2}} = (10\pi)^{\frac{1}{2}} = 5.61$.

UNIVERSITY OF CALIFORNIA.

THE APPROXIMATION OF IRRATIONAL NUMBERS BY FRACTIONS WITH ODD OR EVEN TERMS

By Raphael M. Robinson

Introduction. We shall denote by (μ) the inequality

$$\left|\xi - \frac{A}{B}\right| < \frac{1}{\mu B^2}.$$

Recently, using a geometrical method of proof, W. T. Scott proved the following theorem.[1]

Let ξ be any irrational number, and let any one of the three types of fractions, of the forms odd/odd, odd/even, or even/odd, be selected. Then there are infinitely many fractions A/B of the required type which satisfy (1).

We shall give another proof of this theorem, making use of continued fractions. Furthermore, we shall prove that, *if two of the three types are selected, there are infinitely many fractions of one of these two types which satisfy* (2). These results should be compared with the older theorem of Hurwitz that *the inequality* $(5^{\frac{1}{2}})$ *can be satisfied, if approximations of all three types are allowed.*

Each of these theorems is the best result of its kind; that is, it is not always possible to satisfy (μ) with infinitely many approximations of the required sort, if $\mu > 1$, $\mu > 2$, or $\mu > 5^{\frac{1}{2}}$, respectively. This has been proved by Hurwitz and by Scott for the cases which they considered. For the unrestricted approximations, it is known that only those values of ξ whose continued fraction expansions have partial quotients which beyond a certain point are all equal to 1 do not admit infinitely many approximations satisfying (μ) for some $\mu > 5^{\frac{1}{2}}$, and in fact for $\mu = 2^{\frac{3}{2}}$. For his problem, Scott showed that for any $\mu > 1$ a ξ can be found such that (μ) cannot be satisfied by infinitely many approximations of the required type. We shall show that a ξ independent of μ can be found, and that indeed these exceptional values of ξ have the cardinal number of the continuum. A similar result is obtained in connection with the second problem. In both cases, an exact description of the exceptional irrationals is obtained in terms of the continued fraction expansion.

Finally, we shall solve the same problems in the cases in which we admit approximations of the type even/even in addition to some of the other types.

Continued fractions. We shall state here some results about continued fractions which are well known or easily proved. Let

$$\xi = [q_0, q_1, q_2, \cdots],$$

Received June 1, 1940.

[1] Scott, *Approximation to real irrationals by certain classes of rational fractions*, Bull. Amer. Math. Soc., vol. 46(1940), pp. 124–129.

where the symbol on the right denotes a simple continued fraction. If we put

$$A_{-2} = 0,\ A_{-1} = 1,\ A_n = q_nA_{n-1} + A_{n-2} \qquad (n \geqq 0),$$

$$B_{-2} = 1,\ B_{-1} = 0,\ B_n = q_nB_{n-1} + B_{n-2} \qquad (n \geqq 0),$$

then the convergent $[q_0, q_1, \cdots, q_n]$ is equal to A_n/B_n, and the relation $A_nB_{n-1} - A_{n-1}B_n = \pm 1$ is satisfied. In addition to the convergents, we shall consider as approximations to ξ the fractions

$$\frac{A_n + A_{n-1}}{B_n + B_{n-1}}, \qquad \frac{A_n - A_{n-1}}{B_n - B_{n-1}},$$

which we shall call secondary convergents. Putting

$$\alpha_n = [q_n, q_{n-1}, \cdots, q_1], \qquad \beta_n = [q_{n+1}, q_{n+2}, \cdots],$$

we find that

$$\left|\xi - \frac{A_n}{B_n}\right| = \frac{1}{\lambda_n B_n^2} \qquad \text{with} \quad \lambda_n = \beta_n + \frac{1}{\alpha_n},$$

$$\left|\xi - \frac{A_n + A_{n-1}}{B_n + B_{n-1}}\right| = \frac{1}{\lambda_n'(B_n + B_{n-1})^2} \quad \text{with} \quad \lambda_n' = 1 + \frac{1}{\beta_n - 1} - \frac{1}{\alpha_n + 1},$$

$$\left|\xi - \frac{A_n - A_{n-1}}{B_n - B_{n-1}}\right| = \frac{1}{\lambda_n''(B_n - B_{n-1})^2} \quad \text{with} \quad \lambda_n'' = 1 + \frac{1}{\alpha_n - 1} - \frac{1}{\beta_n + 1}.$$

From the first formula, it is clear that all convergents satisfy (1); noticing that $\lambda_{n-1} = \beta_{n-1} + 1/\alpha_{n-1} = \alpha_n + 1/\beta_n$, hence $\lambda_{n-1} + \lambda_n > 4$, we see that at least one of any two consecutive convergents satisfies (2); and a slightly more complicated argument shows that at least one of any three consecutive convergents satisfies ($5^{1/2}$). Hurwitz's theorem is a consequence of the last statement. Furthermore, it is clear that each secondary convergent satisfies ($\frac{1}{2}$). But the fact which is of particular interest to us is that at least one of the two secondary convergents (for a given value of n) satisfies (1). In fact, the first one does so if $\beta_n < \alpha_n + 2$ and the second if $\alpha_n < \beta_n + 2$.

A partial converse is furnished by the following results. Let A/B be any fraction, and $[r_0, r_1, \cdots, r_n]$ its expansion as a continued fraction. This is seen to be a convergent for all numbers between $[r_0, \cdots, r_{n-1}, r_n + 1]$ and $[r_0, \cdots, r_{n-1}, r_n - \frac{1}{2}]$. Since these numbers are on opposite sides of A/B and differ from it by more than $1/2B^2$, we see that any fraction satisfying (2) is necessarily a convergent to ξ, a result known to Legendre. Similarly, noting that secondary convergents are of the forms $[q_0, q_1, \cdots, q_{n-1}, q_n \pm 1]$, we see that A/B is a convergent or secondary convergent for all numbers between $[r_0, r_1, \cdots, r_{n-1}, r_n + 2]$ and $[r_0, r_1, \cdots, r_{n-1}, r_n - 1]$. These are on opposite sides of A/B and differ from it by more than $1/B^2$, so that any approximation to ξ satisfying (1) is either a convergent or secondary convergent.

Solution of the problem. How does the type of the convergents (odd/odd, odd/even, even/odd) depend on the partial quotients? In the first place, two consecutive convergents cannot be of the same type. For if so, $A_nB_{n-1} - A_{n-1}B_n$ would be even, whereas its value is actually ± 1. If q_n is even, the n-th and $(n-2)$-th convergents are of the same type. If q_n is odd, the n-th convergent is of different type from the $(n-1)$-th and $(n-2)$-th convergents; since there are only three types, this statement is sufficient to determine its type. For convenience, we shall speak of q_n (rather than q_{n-1}) as the partial quotient corresponding to A_{n-1}/B_{n-1}. With this terminology, the convergents on either side of a given convergent are alike or different in type, according as the partial quotient corresponding to the given convergent is even or odd. We notice also that the secondary convergents $(A_n \pm A_{n-1})/(B_n \pm B_{n-1})$ are of type different from A_n/B_n and A_{n-1}/B_{n-1}.

We shall assign to the three types of fractions (odd/odd, odd/even, even/odd) the arbitrary names "round", "square", "curly", in any order. We shall think of the partial quotients $q_0, q_1, q_2, \cdots$ being written in order, and each enclosed in round, square, or curly brackets, according to the type of the corresponding convergent. For example, a segment of this sequence might be (2)[3]{2}[5](1){4}(3). Note that the type of each pair of brackets after the first two is determined by the preceding.

Evidently it is sufficient throughout to consider approximations A/B with $(A, B) = 1$, except when the type even/even is used, in which case we may suppose $(A, B) = 2$. An approximation of type even/even satisfying (μ), when reduced to lowest terms, will satisfy (4μ).

We divide the problem into cases as follows. In case n $(n = 1, 2, 3)$, n of the types odd/odd, odd/even, even/odd are allowed; in case n' $(n = 1, 2, 3, 4)$, n types are allowed, including even/even. Cases 1, 2, 2′, 3′ have three subcases each, depending on which of the types odd/odd, odd/even, even/odd are admitted. However, in describing the exceptional irrationals, we shall not distinguish the subcases.

Case 1. We are to show that the inequality (1) can be satisfied by infinitely many approximations of a prescribed type, say round. Either there are infinitely many round convergents, which satisfy (1); or beyond a certain point the convergents are alternately square and curly. In the latter case, the secondary convergents are round, and at least one of each pair furnishes the required approximation.

What are the exceptional irrationals, which for no $\mu > 1$ admit infinitely many round approximations satisfying (μ)? It is clear that we must have $\lambda_n \to 1$ as $n \to \infty$ through values for which A_n/B_n is round. Hence beyond a certain point the partial quotients in round brackets must be 1, and those on either side must approach ∞. Also, we must have $\max(\lambda_n', \lambda_n'') \to 1$ as $n \to \infty$ through the values for which the n-th and $(n-1)$-th convergents are not round. Since

$$\max\left[\frac{1}{\alpha_n - 1} - \frac{1}{\beta_n + 1}, \frac{1}{\beta_n - 1} - \frac{1}{\alpha_n + 1}\right] > \frac{2}{[\min(\alpha_n, \beta_n)]^2},$$

we see that min $(\alpha_n, \beta_n) \to \infty$, hence $\alpha_n \to \infty$ and $\beta_n \to \infty$, and therefore $q_n \to \infty$ and $q_{n+1} \to \infty$. Thus all partial quotients corresponding to square or curly convergents must approach infinity. It is seen that this condition, together with the condition that all partial quotients corresponding to round convergents are equal to 1, with a finite number of exceptions, is also sufficient.

In order to put this condition into an explicit form, we must take account of the manner in which the type of the convergents depends on the partial quotients. Either the convergents are ultimately all square and curly, in which case the partial quotients are all even beyond a certain point; or else there are infinitely many round convergents. In the latter case the blocks between two consecutive round convergents must be one of certain types, namely,

(1)[even](1), (1)[odd]{odd}(1), (1)[odd]{even}[odd](1),

(1)[odd]{even}[even]{odd}(1), (1)[odd]{even}[even]{even}[odd](1),

or with any longer sequence of alternately square and curly brackets, of which only the first and last enclose odd numbers; or finally, any of these with the square and curly brackets interchanged. The identification of round with one of the types odd/odd, odd/even, even/odd will depend on the early part of the continued fraction. Without going into this, we see that the exceptional irrationals, which for some one of the three types do not admit approximations (μ) with a fixed $\mu > 1$ and arbitrarily large denominator, are those whose continued fraction expansions satisfy the following conditions. The partial quotients, except for the 1's, approach ∞; beyond a certain point, no two 1's are adjacent, the partial quotients next to a 1 but not between two 1's are odd, while the others are even. These exceptional irrationals evidently have the cardinal number of the continuum.

Case 2. We are to show that infinitely many round or square approximations can be found satisfying (2). Now either there are infinitely many pairs of consecutive convergents, of which one is round and the other square, in which case one of each such pair satisfies (2); or beyond a certain point, every other convergent is curly. In the latter case, the intermediate partial quotients must be even, so that the corresponding convergents, which are round or square, satisfy (2).

What are the exceptional irrationals such that for no $\mu > 2$ can (μ) be satisfied by infinitely many round or square approximations? Here we need consider only the convergents, and not the secondary convergents, as approximations. We need only that, for any $\epsilon > 0$, $\lambda_n < 2 + \epsilon$ for all values of n for which the n-th convergent is round or square, with a finite number of exceptions. We notice first that beyond a certain point it is impossible to have three consecutive convergents which are round or square; for (5^1) would be satisfied by one of them. In order to have $\lambda_n < 3$ for the round and square convergents, the corresponding partial quotients must be 1 or 2. Thus the only possible blocks between two consecutive curly convergents are of the forms

$$\{q_n\}(2)\{q_{n+2}\}, \qquad \{q_{n-1}\}(1)[1]\{q_{n+2}\},$$

or the same with the round and square brackets interchanged. In the first case, $\lambda_n \to 2$ gives $q_n \to \infty$ and $q_{n+2} \to \infty$. In the second case, $\lambda_{n-1} + \lambda_n = \alpha_n + 1/\alpha_n + \beta_n + 1/\beta_n$; the condition $\lambda_{n-1} + \lambda_n \to 4$ gives $\alpha_n \to 1$ and $\beta_n \to 1$, from which $q_{n-1} \to \infty$ and $q_{n+2} \to \infty$ follow. Thus the partial quotients in the curly brackets must approach ∞, and this condition is seen to be sufficient. Hence an irrational is exceptional (for one of the three subcases) if and only if its continued fraction expansion has partial quotients obtained by taking a sequence of integers which tend to infinity, and from a certain point on inserting either a 2 or two 1's between consecutive terms. This set has the cardinal number of the continuum.

Case 3. This is the problem solved by Hurwitz. The results are stated in the introduction. Unlike Cases 1 and 2, here the exceptional irrationals are denumerable.

Case 1′. Here only fractions even/even are permitted. If the reduced form of the fractions is considered, we have Case 3. Hence the best inequality which can be satisfied is $(5^{\frac{1}{2}}/4)$, and the exceptional irrationals are the same as there.

Case 2′. Here we permit approximations even/even and one other type, say round. We shall see that the best inequality is $(5^{\frac{1}{2}}/2)$. For if there are infinitely many partial quotients greater than 4, then (5) can be satisfied by infinitely many approximations, or (5/4) by infinitely many approximations even/even. Otherwise, beyond a certain point, $\lambda_n > 1 + 1/5 + 1/5$, so that if there are infinitely many round convergents, then (7/5) can be satisfied by infinitely many round approximations. In the remaining case the convergents are alternately square and curly, so that all partial quotients are even, hence 2 or 4. The secondary convergents are round. If $q_{n+1} = 2$ then $\beta_n < 3$, $\alpha_n > 2$, hence $\lambda_n' > 7/6$. Thus at least (7/6) can be satisfied in all cases except the one in which all partial quotients beyond a certain point are equal to 4. In this case, $(5^{\frac{1}{2}}/2)$ is the best inequality which can be satisfied; provided that the type of convergent of which there are but a finite number is called round. The exceptional irrationals for the three subcases together are those whose continued fraction expansions have partial quotients which are all 4's beyond a certain point; these are denumerable.

Also in this case there is a relation to Case 3, which we shall indicate for one of the three subcases, in which the permitted type is even/odd. Then all approximations $2A/B$ are permitted; or we wish to approximate $\xi/2$ by A/B. This argument leads to the answer $(5^{\frac{1}{2}}/2)$, and shows that the exceptional irrationals (for this subcase) are those obtained by doubling irrationals whose continued fraction expansions end with 1's.

Case 3′. Here we admit all approximations except one of the types odd/odd, odd/even, or even/odd; say curly is the type excluded. This is the only one of the cases admitting even/even which cannot be reduced to Case 3 in some way; here the answer involves a new quantity, the best inequality being $(65^{\frac{1}{2}}/4)$. To satisfy $(65^{\frac{1}{2}}/4)$ by a fraction even/even is the same as satisfying $(65^{\frac{1}{2}})$ by a reduced fraction. If the desired inequality is impossible, no partial quotient

can exceed 8, and those corresponding to round or square convergents cannot exceed 2. If either of these bounds were attained, we should have $\lambda_n > 8 + 1/3 + 1/3 > 65^{1/2}$ or $\lambda_n > 2 + 1/9 + 1/9 > 65^{1/2}/4$, respectively; hence the bounds can be reduced to 7 and 1. That is, all partial quotients corresponding to round or square convergents must be equal to 1, and the others must not exceed 7. Thus between successive curly convergents we can have only a sequence of the form $\{q_{n-1}\}(1)[1]\{q_{n+2}\}$, or the same with the round and square brackets interchanged. If now $q_{n-1} < 7$, then $\alpha_n > 8/7$, and hence $\alpha_n + 1/\alpha_n > 8/7 + 7/8$; and in any case $\beta_n > 9/8$, and hence $\beta_n + 1/\beta_n > 9/8 + 8/9$. Thus $\lambda_{n-1} + \lambda_n > 8/7 + 7/8 + 9/8 + 8/9 = 4 + 2/63$, so that $\max(\lambda_{n-1}, \lambda_n) > 2 + 1/63 > 65^{1/2}/4$. Hence except for continued fractions ending 7, 1, 1, 7, 1, 1, 7, 1, 1, 7, $\cdots$ an even stronger inequality can be satisfied. In this case (for a suitable meaning of curly), $(65^{1/2}/4)$ is seen to be the best inequality which can be satisfied; the exceptional irrationals are denumerable.

Case 4′. The result here is evidently the same as in Case 3.

UNIVERSITY OF CALIFORNIA.

THE JUMP OF ALMOST PERIODIC FUNCTIONS AND OF FOURIER INTEGRALS

BY OTTO SZÁSZ

1. We have proved recently [2][1] the following:

THEOREM A. *Suppose that $f(t)$ is integrable L in $(-\pi, \pi)$, and has period 2π; if for a fixed x there exists a $D(x)$ such that*

$$\int_0^h |f(x+t) - f(x-t) - D(x)|\, dt = O(h) \quad \text{as } h \downarrow 0 \tag{1}$$

and

$$\int_0^h \{f(x+t) - f(x-t) - D(x)\}\, dt = o(h) \quad \text{as } h \downarrow 0, \tag{2}$$

then

$$\frac{1}{2n}\left\{\sum_{\nu=1}^{n} \nu B_\nu(x) + \sum_{\nu=n+1}^{2n} (2n-\nu) B_\nu(x)\right\} \to \frac{D(x)\log 2}{\pi} \quad \text{as } n \uparrow \infty;$$

here

$$\sum_{\nu=1}^{n} B_\nu(t) = \sum_{1}^{n} (b_\nu \cos \nu t - a_\nu \sin \nu t) \qquad (n = 1, 2, 3, \cdots)$$

are the partial sums of the conjugate Fourier series corresponding to $f(t)$.

$D(x)$ is the generalized jump of $f(t)$ at a point x; if in particular $f(x+t) - f(x-t) \to D(x)$ as $t \downarrow 0$, then (1) and (2) obviously hold.

We shall give here analogous results for generalized Fourier series and for Fourier integrals.

2. We consider real-valued almost-periodic functions in the sense of Besicovitch ([1], Chapter II). If $f(t)$ is a B.a.p. function, then

$$M\{f(t)e^{-i\lambda t}\} = \lim_{h\uparrow\infty} \frac{1}{2h}\int_{-h}^{h} f(t)e^{-i\lambda t}\, dt = c(\lambda)$$

exists for all real values of λ, and it may differ from zero for at most an enumerable set of values λ ([1], Chapter II, §8). Of these let the positive λ be arranged in some order:

$$\lambda_1, \lambda_2, \lambda_3, \cdots, \lambda_n > 0,$$

Received June 11, 1940; presented to the American Mathematical Society, September 12, 1940.

[1] Numbers in brackets refer to the bibliography at the end of this paper.

and let

$$c(0) = a_0, \qquad c(\lambda_\nu) = \tfrac{1}{2}(a_\nu - ib_\nu);$$

then

$$f(t) \sim a_0 + \sum_{\nu=1}^{\infty} (a_\nu \cos \lambda_\nu t + b_\nu \sin \lambda_\nu t) \equiv \sum_{\nu=0}^{\infty} A_\nu(t)$$

is the generalized Fourier series of $f(t)$.

We then have

$$f(x + t) - f(x - t) \sim 2 \sum_{1}^{\infty} (b_\nu \cos \lambda_\nu x - a_\nu \sin \lambda_\nu x) \sin \lambda_\nu t \equiv 2 \sum_{1}^{\infty} \bar{A}_\nu(x) \sin \lambda_\nu t.$$

On putting

$$f(x + t) - f(x - t) = 2\psi(t)$$

we have $\psi(-t) = -\psi(t)$, and

$$\psi(t) \sim \sum_{1}^{\infty} \bar{A}_\nu(x) \sin \lambda_\nu t,$$

where

$$\bar{A}_\nu(x) = \lim_{h \uparrow \infty} \frac{2}{h} \int_0^h \psi(t) \sin \lambda_\nu t \, dt.$$

Thus, to determine the jump of a function $f(t)$ at a given point $t = x$, we may assume without loss of generality that the given point is $t = 0$, and that the function is an odd function: $\psi(-t) = -\psi(t)$. Its Fourier series has the form

$$\psi(t) \sim \sum_{1}^{\infty} b_\nu \sin \lambda_\nu t, \qquad \lambda_\nu > 0, \tag{3}$$

where

$$b_n = \lim_{h \uparrow \infty} \frac{2}{h} \int_0^h \psi(t) \sin \lambda_n t \, dt \qquad (n = 1, 2, 3, \ldots).$$

In the proof of Theorem A we have used the formula

$$\sum_{1}^{n} \nu B_\nu + \sum_{n+1}^{2n} (2n - \nu) B_\nu = \frac{2}{\pi} \int_0^\pi \psi(t) \sin nt \left(\frac{\sin \frac{1}{2} nt}{\sin \frac{1}{2} t} \right)^2 dt.$$

Accordingly we introduce here

$$P(\eta) = \frac{4}{\pi} \int_0^\infty \psi(t) \sin 2\eta t \, \frac{\sin^2 \eta t}{\eta t^2} \, dt. \tag{4}$$

The absolute convergence of this integral follows from

$$\overline{\lim_{h \uparrow \infty}} \frac{1}{h} \int_0^h |\psi(t)| \, dt < \infty, \tag{5}$$

and from the ([5], p. 138)

Lemma. *If* (5) *holds, then*

$$\int_0^\infty (1 + t^2)^{-1} \mid \psi(t) \mid dt < \infty.$$

In fact, writing $\int_0^a \mid \psi(t) \mid dt = v(a)$, we find for $a > 0$

$$\int_0^a (1 + t^2)^{-1} \mid \psi(t) \mid dt = \frac{v(a)}{1 + a^2} + 2 \int_0^a (1 + t^2)^{-2} t\, v(t)\, dt$$

$$= o(1) + O\left(\int_0^a t^2(1 + t^2)^{-2}\, dt\right) = O(1) \quad \text{as } a \uparrow \infty.$$

This proves the lemma and the absolute convergence of (4). Let us use the formula

$$\frac{2}{\pi} \int_0^\infty \cos 2kt \frac{\sin^2 \eta t}{\eta t^2} dt = \begin{cases} 1 - \dfrac{\mid k \mid}{\eta} & \text{if} \quad \mid k \mid < \eta, \\ 0 & \text{if} \quad \mid k \mid \geqq \eta, \end{cases}$$

and the development (3); formal termwise integration then yields

$$P(\eta) \sim \frac{4}{\pi} \sum_1^\infty b_\nu \int_0^\infty \sin \lambda_\nu t \sin 2\eta t \frac{\sin^2 \eta t}{\eta t^2} dt$$

$$\sim \frac{2}{\pi} \sum_1^\infty b_\nu \int_0^\infty \{\cos (\lambda_\nu - 2\eta)t - \cos (\lambda_\nu + 2\eta)t\} \frac{\sin^2 \eta t}{\eta t^2} dt$$

$$\sim \frac{1}{2\eta} \sum_{0 < \lambda_n \leqq 2\eta} \lambda_n b_n + \sum_{2\eta < \lambda_n < 4\eta} \left(2 - \frac{\lambda_n}{2\eta}\right) b_n.$$

This equivalence becomes an equality in special cases. The connection of the expression (4) with the jump is given by

Theorem 1. *Let $\psi(t)$ be a function satisfying* (5); *assume that for a certain D*

$$\int_0^h \mid \psi(t) - D \mid dt = o(h) \quad \text{as } h \downarrow 0.$$

Then

$$\frac{\pi}{4} P(\eta) \equiv \int_0^\infty \psi(t) \sin 2\eta t \frac{\sin^2 \eta t}{\eta t^2} dt \to D \log 2 \quad \text{as } \eta \uparrow \infty.$$

We first evaluate the integral

$$I = \int_0^\infty \sin 2\eta t \frac{\sin^2 \eta t}{\eta t^2} dt = \int_0^\infty \sin 2\tau \left(\frac{\sin \tau}{\tau}\right)^2 d\tau.$$

This can be done by using complex integration. However, the following way is simpler. Evidently

$$I = \tfrac{1}{2} \int_0^\infty t^{-2}(\sin 2t - \tfrac{1}{2} \sin 4t)\, dt = \tfrac{1}{2} \lim_{\epsilon \downarrow 0} \int_\epsilon^\infty \equiv \tfrac{1}{2} \lim_{\epsilon \downarrow 0} I(\epsilon).$$

Now

$$\tfrac{1}{2}I(\epsilon) = \int_{2\epsilon}^{\infty} \tau^{-2} \sin \tau \, d\tau - \int_{4\epsilon}^{\infty} \tau^{-2} \sin \tau \, d\tau = \int_{2\epsilon}^{4\epsilon} \tau^{-2} \sin \tau \, d\tau$$

$$= \int_{2}^{4} \frac{\sin \epsilon u}{\epsilon u} u^{-1} du = \int_{2}^{4} u^{-1} du - \int_{2}^{4} \left(1 - \frac{\sin \epsilon u}{\epsilon u}\right) u^{-1} du$$

$$\to \log 2 \quad \text{as } \epsilon \downarrow 0.$$

Hence $I = \log 2$.

We now consider

$$\frac{\pi}{4} P(\eta) - D \log 2 = \int_0^{\infty} \{\psi(t) - D\} \sin 2\eta t \frac{\sin^2 \eta t}{\eta t^2} dt, \tag{6}$$

which yields

$$\left| \frac{\pi}{4} P(\eta) - D \log 2 \right| \leqq \int_0^{\infty} | \psi(t) - D | \frac{\sin^2 \eta t}{\eta t^2} dt.$$

From the elementary inequality

$$(1 + x) | \sin x | < 2x \qquad \text{for } x > 0$$

we get for $x = \eta t$

$$\frac{\sin^2 \eta t}{\eta t^2} < \frac{4\eta}{(1 + \eta t)^2}, \qquad \eta > 0,\ t > 0. \tag{7}$$

Thus

$$\left| \frac{\pi}{4} P(\eta) - D \log 2 \right| \leqq 4\eta \int_0^{\infty} | \psi(t) - D | (1 + \eta t)^{-2} dt.$$

If we write $\int_0^t | \psi(\tau) - D | \, d\tau = u(t)$, integration by parts gives

$$\left| \frac{\pi}{4} P(\eta) - D \log 2 \right| \leqq 4\eta u(t)(1 + \eta t)^{-2} \Big]_0^{\infty} + 8\eta^2 \int_0^{\infty} u(t)(1 + \eta t)^{-3} dt.$$

Let

$$\epsilon(t) = \max_{0 < \tau \leqq t} \frac{u(\tau)}{\tau} < \gamma, \qquad t > 0,\ \gamma \text{ a constant.}$$

Then, by hypothesis, $\epsilon(t) \downarrow 0$ as $t \downarrow 0$. Also

$$\left| \frac{\pi}{4} P(\eta) - D \log 2 \right| \leqq 8\eta^2 \int_0^{\infty} t\,\epsilon(t)(1 + \eta t)^{-3} dt$$

$$\leqq 8\eta \int_0^{\infty} \epsilon(t)(1 + \eta t)^{-2} dt = 8\eta \left(\int_0^{\eta^{-1}} + \int_{\eta^{-1}}^{\infty} \right)$$

$$\equiv C_1(\eta) + C_2(\eta).$$

Now

$$C_2(\eta) < 8\eta\gamma \int_{\eta^{-\frac{1}{2}}}^{\infty} (1 + \eta t)^{-2}\, dt = 8\gamma \Big[-(1 + \eta t)^{-1} \Big]_{\eta^{-\frac{1}{2}}}^{\infty} = \frac{8\gamma}{1 + \eta^{\frac{1}{2}}}$$

and

$$C_1(\eta) < 8\eta\epsilon(\eta^{-\frac{1}{2}}) \int_0^{\infty} (1 + \eta t)^{-2}\, dt = 8\epsilon(\eta^{-\frac{1}{2}}).$$

Thus, letting $\eta \uparrow \infty$, we get Theorem 1. For a similar result cf. [5], p. 145.

Next we prove the more general

THEOREM 2. *Let $\psi(t)$ be the same as in Theorem* 1. *Assume*

$$u(h) \equiv \int_0^h |\psi(t) - D|\, dt = O(h) \quad \textit{as } h \downarrow 0$$

and

$$w(h) \equiv \int_0^h \{\psi(t) - D\}\, dt = o(h) \quad \textit{as } h \downarrow 0.$$

Then again $\frac{1}{4}\pi P(\eta) \to D \log 2$ as $\eta \uparrow \infty$.

For the proof let

$$\max_{0<h\leqq t} \frac{|w(h)|}{h} = \delta(t) \downarrow 0 \quad \text{as } t \downarrow 0.$$

Hence for any $\eta > 0$ the equation

$$\rho\delta(\rho)^{\frac{1}{2}} = \frac{1}{\eta}$$

has a unique solution: $\rho = \rho(\eta) \downarrow 0$ as $\eta \uparrow \infty$. Let now

$$(8) \qquad \int_0^{\infty} \{\psi(t) - D\} \sin 2\eta t \,\frac{\sin^2 \eta t}{\eta t^2}\, dt = \int_0^{\rho(\eta)} + \int_{\rho(\eta)}^{\infty} \equiv K_1(\eta) + K_2(\eta).$$

Then, if (7) is used again, integration by parts yields

$$|K_2(\eta)| \leqq 8\eta^2 \int_{\rho}^{\infty} u(t)(1 + \eta t)^{-3}\, dt < 8\eta\gamma \int_{\rho}^{\infty} (1 + \eta t)^{-2}\, dt = \frac{8\gamma}{1 + \eta\rho}.$$

But $\eta\rho = \delta^{-\frac{1}{2}} \uparrow \infty$ as $\eta \uparrow \infty$. Hence $K_2(\eta) \to 0$. Furthermore

$$K_1(\eta) = \frac{1}{2\eta} \int_0^{\rho} \{\psi(t) - D\}(\sin 2\eta t - \tfrac{1}{2} \sin 4\eta t)t^{-2}\, dt$$

or

$$\eta K_1(\tfrac{1}{2}\eta) = \int_0^{\rho} \{\psi(t) - D\}(\sin \eta t - \tfrac{1}{2} \sin 2\eta t)t^{-2}\, dt.$$

Now, integration by parts yields

$$\eta K_1(\tfrac{1}{2}\eta) = w(\rho)(\sin \eta\rho - \tfrac{1}{2} \sin 2\eta\rho)\rho^{-2}$$

$$- \int_0^\rho w(t)\{(\cos \eta t - \cos 2\eta t)\eta t^{-2} - 2(\sin \eta t - \tfrac{1}{2} \sin 2\eta t)t^{-3}\}\, dt$$

$$= w(\rho) \sin \eta\rho\, (1 - \cos \eta\rho)\rho^{-2} - \eta \int_0^\rho w(t)(\cos \eta t - 1 + 2 \sin^2 \eta t)t^{-2}\, dt$$

$$+ 2 \int_0^\rho w(t) \sin \eta t\, (1 - \cos \eta t)t^{-3}\, dt$$

$$= 2w(\rho) \sin \eta\rho \sin^2 (\tfrac{1}{2}\eta\rho)\, \rho^{-2} - 2\eta \int_0^\rho w(t)\{\sin^2 \eta t - \sin^2(\tfrac{1}{2}\eta t)\}t^{-2}\, dt$$

$$+ 4 \int_0^\rho w(t) \sin \eta t \sin^2 (\tfrac{1}{2}\eta t)\, t^{-3}\, dt.$$

Hence

$$| K_1(\tfrac{1}{2}\eta) | < 2\, \frac{w(\rho)}{\rho}\, (\eta\rho)^{-1} + 10\delta(\rho) \int_0^\rho t^{-1} \sin^2 (\tfrac{1}{2}\eta t)\, dt$$

$$+ 8\delta(\rho)\eta^{-1} \int_0^\rho t^{-2} \sin^2 (\tfrac{1}{2}\eta t)\, dt$$

$$< 2\delta^{\frac{3}{4}} + \delta\eta^2\rho^2 + \delta\eta^2\rho^2 = 2\delta^{\frac{3}{4}} + 2\delta^{\frac{1}{2}} \to 0 \quad \text{as } \eta \to \infty.$$

This proves the theorem (cf. (6) and (8)).

3. Next we consider Fourier integrals. Without loss of generality we may restrict ourselves to sine-transforms and to the point $t = 0$.

Let $g(t)$ be in $L^p(0, \infty)$ $(1 \leqq p \leqq 2)$; let $g(-t) = -g(t)$ and let $p' = p/(p - 1)$ for $p > 1$. If $G(t)$ is the transform of $g(t)$, then ([4], p. 96)

$$g(x) = \left(\frac{2}{\pi}\right)^{\frac{1}{2}} \frac{d}{dx} \int_0^\infty G(t)\, \frac{1 - \cos tx}{t}\, dt$$

almost everywhere. We shall prove the following

THEOREM 3. *Let $g(t)$ belong to $L^p(0, \infty)$ $(1 \leqq p \leqq 2)$, $g(-t) = -g(t)$ and let $G(t)$ be its sine-transform. Let*

$$V(\lambda) = \int_0^\lambda \left(1 - \frac{t}{\lambda}\right) G(t)\, dt, \qquad \lambda > 0,$$

and suppose that for a constant d

$$\varphi(h) \equiv \int_0^h | g(t) - d |\, dt = O(h) \quad \text{as } h \downarrow 0$$

and

$$\Phi(h) \equiv \int_0^h \{g(t) - d\}\, dt = o(h) \quad \textit{as } h \downarrow 0.$$

Then

$$V(2\lambda) - V(\lambda) \to d\left(\frac{2}{\pi}\right)^{\frac{1}{2}} \log 2 \quad \textit{as } \lambda \to \infty.$$

For the proof we use the formula ([3], p. 174)

$$V(2\lambda) - V(\lambda) = \left(\frac{2}{\pi}\right)^{\frac{1}{2}} \int_0^\infty g(t) \sin \lambda t \frac{1 - \cos \lambda t}{\lambda t^2}\, dt$$

and (cf. §2)

$$I = \int_0^\infty \sin \lambda t \frac{1 - \cos \lambda t}{\lambda t^2}\, dt = \log 2.$$

Then

$$V(2\lambda) - V(\lambda) - d\left(\frac{2}{\pi}\right)^{\frac{1}{2}} \log 2 = \left(\frac{2}{\pi}\right)^{\frac{1}{2}} \int_0^\infty \{g(t) - d\} \sin \lambda t \frac{1 - \cos \lambda t}{\lambda t^2}\, dt.$$

For the rest of the proof we now refer to the previous theorem.

Bibliography

1. A. S. Besicovitch, *Almost periodic functions*, Cambridge, 1932.
2. Otto Szász, *The jump of a function determined by its Fourier coefficients*, this Journal, vol. 4(1938), pp. 401–407.
3. Otto Szász, *Some extremum problems in the theory of Fourier series*, American Journal of Mathematics, vol. 61(1939), pp. 165–177.
4. E. C. Titchmarsh, *Introduction to the Theory of Fourier Integrals*, Oxford, 1937.
5. N. Wiener, *The Fourier Integral and Certain of Its Applications*, Cambridge, 1933.

University of Cincinnati.

THE BINARY POLYHEDRAL GROUPS, AND OTHER GENERALIZATIONS OF THE QUATERNION GROUP

By H. S. M. Coxeter

1. **Introduction.** Hamilton's formulas

$$i^2 = j^2 = k^2 = ijk = -1$$

suggest the following definition for the quaternion group:

$$R^2 = S^2 = T^2 = RST \neq 1.$$

The natural generalization is

(1.1) $$R^l = S^m = T^n = RST.$$

Let $\langle l, m, n\rangle$ denote the (largest) group defined by (1.1). This is symmetrical among l, m, n: for cyclic permutation, obviously; and for transposition, by changing R, S, T into T^{-1}, S^{-1}, R^{-1}, respectively.

Any two of R, S, T suffice to generate $\langle l, m, n\rangle$. For, if

(1.2) $$R^l = S^m = T^n = RST = Z,$$

we can substitute $ZT^{-1}S^{-1}$ for R, obtaining

(1.3) $$S^m = T^n = Z, \qquad (ST)^l = Z^{l-1}.$$

In particular, $\langle 2, m, n\rangle$ is simply defined by

(1.4) $$S^m = T^n = (ST)^2.$$

Another definition ıor $\langle 2, m, n\rangle$ comes from the observation that $R = ST$. Substituting $S^{-1}R$ for T in (1.1), we obtain $R^2 = S^m = (S^{-1}R)^n$, or, writing S^{-1} for S,

(1.5) $$R^2 = S^{-m} = (RS)^n.$$

In particular, $\langle 2, 2, m\rangle$ is the same group as $\langle 2, 2, -m\rangle$.

The relations (1.4) and (1.5) are reminiscent of Miller's[1]

$$s_1^m = s_2^n, \qquad (s_1s_2)^2 = 1$$

and

$$s_1^2 = s_2^n, \qquad (s_1s_2)^l = 1,$$

but are by no means identical with them.

Received June 20, 1940; presented to the American Mathematical Society, April 26, 1940.

[1] G. A. Miller, *Generalization of the groups of genus zero*, Transactions of the American Mathematical Society, vol. 8(1907), pp. 1–13.

W. Threlfall has observed[2] that

$$(1.6) \qquad \langle 2, 2, n\rangle, \quad \langle 2, 3, 3\rangle, \quad \langle 2, 3, 4\rangle, \quad \langle 2, 3, 5\rangle$$

are the *binary polyhedral groups*, of order

$$4(l^{-1} + m^{-1} + n^{-1} - 1)^{-1}.$$

After giving a new proof of this result, we shall show that, except in the special cases $\langle -2, 2, n\rangle$ (n odd) and $\langle \pm 2, -3, 3\rangle$, each of the groups

$$(1.7) \quad \langle \pm 2, \pm 2, n\rangle, \quad \langle \pm 2, \pm 3, \pm 3\rangle, \quad \langle \pm 2, \pm 3, \pm 4\rangle, \quad \langle \pm 2, \pm 3, \pm 5\rangle$$

is the direct product of the corresponding one of (1.6) with the cyclic group of order

$$| l^{-1} + m^{-1} + n^{-1} - 1 | (| l |^{-1} + | m |^{-1} + | n |^{-1} - 1)^{-1}.$$

Moreover, even in the exceptional cases, it will be seen that the order of $\langle l, m, n\rangle$ is still

$$4 | l^{-1} + m^{-1} + n^{-1} - 1 | (| l |^{-1} + | m |^{-1} + | n |^{-1} - 1)^{-2}.$$

These results show, in particular, that the eight groups

$$\langle 2, 3, 5\rangle, \quad \langle 2, 3, -5\rangle, \quad \langle 2, -3, 5\rangle, \quad \langle -2, 3, 5\rangle,$$
$$\langle 2, -3, -5\rangle, \quad \langle -2, 3, -5\rangle, \quad \langle -2, -3, 5\rangle, \quad \langle -2, -3, -5\rangle$$

are Hölder's[3]

$$\mathfrak{K}_{60}^{(2)}, \quad \mathfrak{K}_{60}^{(22)}, \quad \mathfrak{K}_{60}^{(38)}, \quad \mathfrak{K}_{60}^{(58)}, \quad \mathfrak{K}_{60}^{(62)}, \quad \mathfrak{K}_{60}^{(82)}, \quad \mathfrak{K}_{60}^{(98)}, \quad \mathfrak{K}_{60}^{(122)},$$

respectively. The groups $\langle 2, 3, n\rangle$, for $n = -3, -4, -5$, have already been investigated by M. Dehn and H. Seifert.[2]

2. Trivial cases. We exclude from consideration the cases where the defining relations (1.1) necessitate more or less arbitrary conventions before they become significant, namely, the cases where $lmn = 0$, or $l = 1$ and $m + n = 0$, etc. When $l = 1$, we have $S^m = T^n$, $ST = 1$; therefore

$$\langle 1, m, n\rangle \text{ is the cyclic group of order } | m + n | \qquad (m + n \neq 0).$$

When $l = -1$, (1.3) gives $ST = Z^2 = S^{2m}$, whence $T = S^{2m-1}$, $S^m = T^n = S^{(2m-1)n}$, and $S^{2mn-m-n} = 1$; therefore

$$\langle -1, m, n\rangle \text{ is the cyclic group of order } | 2mn - m - n | \quad (mn \neq 1).$$

Having disposed of these, we assume from now on that $| l |$, $| m |$, $| n |$ are all greater than 1.

[2] Jahresbericht der deutschen Mathematiker-Vereinigung, vol. 41(1932), pp. *6–8*; vol. 46(1936), p. *80*. (The present paper arose from an attempt to solve Aufgabe 235.)

[3] O. Hölder, *Bildung zusammengesetzter Gruppen*, Mathematische Annalen, vol. 46(1895), pp. 321–422; p. 354.

3. **The polyhedral factor group.** The element Z of (1.2), being permutable with R, S, T, generates an invariant subgroup whose quotient group is the polyhedral group (l, m, n), defined by

$$R^l = S^m = T^n = RST = 1.$$

Hence the order of $\langle l, m, n\rangle$ is g times the period of Z, where g is the order of (l, m, n) or $(|l|, |m|, |n|)$.

If l, m, n are greater than 1, we know that the group (l, m, n) is finite when

$$l^{-1} + m^{-1} + n^{-1} > 1$$

and infinite otherwise. In the finite case its elements can be represented by rotations about concurrent lines in ordinary space; the fundamental region consists of two spherical triangles, each having angles π/l, π/m, π/n. Therefore $2g$ such triangles cover their sphere, and

$$g = 2(l^{-1} + m^{-1} + n^{-1} - 1)^{-1}.$$

Since $(|l|, |m|, |n|)$ is a factor group of $\langle l, m, n\rangle$, the latter is infinite whenever

$$|l|^{-1} + |m|^{-1} + |n|^{-1} \leqq 1.$$

Thus the finite groups that remain to be investigated are just (1.7), and our main problem is to prove that in these cases the period of Z is

$$|l^{-1} + m^{-1} + n^{-1} - 1|\cdot g,$$

where

$$g = 2(|l|^{-1} + |m|^{-1} + |n|^{-1} - 1)^{-1}.$$

4. **The case when l, m, n are positive.** There is no loss of generality in assuming that $l \leqq m \leqq n$. Since the inequalities

$$l > 1, \quad m > 1, \quad n > 1, \quad l^{-1} + m^{-1} + n^{-1} > 1$$

then imply $l = 2$, what we have to show is that the relations

$$S^m = T^n = (ST)^2 = Z \tag{4.1}$$

imply $Z^2 = 1$ in the following cases:

(a) $m = 2$; (b) $m = n = 3$;

(c) $m = 3, n = 4$; (d) $m = 3, n = 5$.

(a) When $m = 2$, we have the relations

$$S^2 = T^n = (ST)^2,$$

which imply $T = ST^{-1}S^{-1}$, whence

$$Z = S^2 = T^n = (ST^{-1}S^{-1})^n = ST^{-n}S^{-1} = SZ^{-1}S^{-1} = Z^{-1}.$$

(b) When $m = 3$, we have the relations

$$S^3 = T^n = (ST)^2,$$

which imply $T = S^2T^{-1}S^{-1}$, $S = T^{-1}S^{-1}T^{n-1}$, whence

$$\begin{aligned} Z = S^3 = T^n &= (S^2T^{-1}S^{-1})^n = S(ST^{-1})^nS^{-1} \\ &= (ST^{-1})^n \qquad (4.2) \\ &= (T^{-1}S^{-1}T^{n-2})^n = T^{-1}(S^{-1}T^{n-3})^nT \\ &= (S^{-1}T^{n-3})^n. \qquad (4.3) \end{aligned}$$

When $n = 3$, this gives at once

$$Z = S^{-3} = Z^{-1}.$$

(c) When $n = 4$, (4.3) and (4.2) give

$$Z = (S^{-1}T)^4 = (ST^{-1})^{-4} = Z^{-1}.$$

(d) When $n = 5$, (4.3) gives

$$\begin{aligned} Z = (S^{-1}T^2)^5 &= \{S^{-1}(S^2T^{-1}S^{-1})^2\}^5 \\ &= (ST^{-1}ST^{-1}S^{-1})^5 = ST(T^{-2}S)^5T^{-1}S^{-1} \\ &= (S^{-1}T^2)^{-5} = Z^{-1}. \end{aligned}$$

(We shall see, in our final section, that when $n = 6$ the period of Z is unrestricted.)

5. Representation by quaternions. We have proved that the relations (4.1) imply $Z^2 = 1$; but before we can assert that Z *is of period two*, we must show that they do not imply $Z = 1$. This can be done in various ways.[4] One way is to represent the elements of each group by quaternions, and verify that Z is represented by the quaternion -1.

(a) This is a natural generalization of the representation

$$R = i, \quad S = j, \quad T = k$$

for $\langle 2, 2, 2\rangle$. We observe that the relations

$$R^2 = S^2 = T^n = RST = -1,$$

which define $\langle 2, 2, n\rangle$, are satisfied by

$$R = i, \quad S = i\cos(\pi/n) + j\sin(\pi/n), \quad T = \cos(\pi/n) + k\sin(\pi/n).$$

(b) For $\langle 2, 3, 3\rangle$, we have

$$R = i, \quad S = \tfrac{1}{2}(1 + i + j - k), \quad T = \tfrac{1}{2}(1 + i + j + k).$$

[4] Jahresbericht der deutschen Mathematiker-Vereinigung, vol. 41(1932), pp. *6–8*; vol. 42(1933), p. *3*; vol. 47(1937), p. *42*.

(c) For $\langle 2, 3, 4\rangle$, we have

$$R = 2^{-\frac{1}{2}}(i + j), \quad S = \tfrac{1}{2}(1 + i + j + k), \quad T = 2^{-\frac{1}{2}}(1 + i).$$

(d) For $\langle 2, 3, 5\rangle$, we have

$$R = i, \quad S = \tfrac{1}{2}(\tau + i + \tau^{-1}j), \quad T = \tfrac{1}{2}(1 + \tau i + \tau^{-1}k),$$

where

$$\tau = \tfrac{1}{2}(5^{\frac{1}{2}} + 1) = 2 \cos (\tfrac{1}{5}\pi).$$

These results show that, whenever l, m, n and $l^{-1} + m^{-1} + n^{-1}$ are all greater than 1, the period of Z in $\langle l, m, n\rangle$ is 2, and consequently the order of $\langle l, m, n\rangle$ is

$$2g = 4(l^{-1} + m^{-1} + n^{-1} - 1)^{-1}.$$

6. **Further remarks on the binary polyhedral groups.** The choice of the above quaternions was determined by geometrical considerations. Cayley[5] showed that, in Riemann's representation of the complex numbers $z = x + \iota y$ by the points of a sphere of unit radius, a rotation through angle 2θ about the line joining points (ξ, η, ζ), $(-\xi, -\eta, -\zeta)$ represents the linear fractional transformation

$$z' = \frac{Az - \bar{C}}{Cz + \bar{A}},$$

where

$$A = \cos \theta + \iota\zeta \sin \theta, \qquad C = (\eta + \iota\xi) \sin \theta.$$

Moreover,[6] the corresponding *binary* transformation

$$z_1' = Az_1 - \bar{C}z_2, \qquad z_2' = Cz_1 + \bar{A}z_2$$

can be represented by the quaternion

$$\cos \theta + (\xi i + \eta j + \zeta k) \sin \theta,$$

in such a way that the product of two such transformations is represented by the product of the corresponding quaternions.[7] Thus the identity and the rotations through π about the Cartesian axes correspond to the quaternion units ± 1, $\pm i$, $\pm j$, $\pm k$, and any rotation group of order g corresponds to a group of

[5] A. Cayley, *On the correspondence of homographies and rotations*, Mathematische Annalen, vol. 15(1879), pp. 238–240.

[6] A. Cayley, *On certain results relating to quaternions*, Philosophical Magazine, (3), vol. 26(1845), pp. 141–145; G. Boole, *Notes on quaternions*, ibid., vol. 33(1848), pp. 278–280; F. Klein, *Vorlesungen über das Ikosaeder*, Leipzig, 1884, pp. 35–36.

[7] This is done by writing 1, i, j, k for the matrices

$$\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \quad \begin{pmatrix} 0 & \iota \\ \iota & 0 \end{pmatrix}, \quad \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}, \quad \begin{pmatrix} \iota & 0 \\ 0 & -\iota \end{pmatrix}.$$

quaternions of order $2g$. In this manner, the dihedral group leads to the *dicyclic* group $\langle 2, 2, n\rangle$, whose $4n$ elements are

$$(6.1)\qquad i\cos(r\pi/n) + j\sin(r\pi/n), \qquad \cos(r\pi/n) + k\sin(r\pi/n) \qquad (r = 0, 1, \cdots, 2n-1);$$

the tetrahedral group leads to the *binary tetrahedral* group $\langle 2, 3, 3\rangle$, whose 24 elements are[8]

$$(6.2)\qquad \pm 1, \quad \pm i, \quad \pm j, \quad \pm k, \quad \tfrac{1}{2}(\pm 1 \pm i \pm j \pm k);$$

the octahedral group leads to the *binary octahedral* group $\langle 2, 3, 4\rangle$, whose 48 elements are (6.2) together with[8]

$$(6.3)\qquad \begin{cases} (\pm 1 \pm i)2^{-\frac{1}{2}}, & (\pm 1 \pm j)2^{-\frac{1}{2}}, & (\pm 1 \pm k)2^{-\frac{1}{2}}, \\ (\pm j \pm k)2^{-\frac{1}{2}}, & (\pm k \pm i)2^{-\frac{1}{2}}, & (\pm i \pm j)2^{-\frac{1}{2}}; \end{cases}$$

and, finally, the icosahedral group leads to the *binary icosahedral* group $\langle 2, 3, 5\rangle$, whose 120 elements are (6.2) together with

$$(6.4)\qquad \begin{cases} \tfrac{1}{2}(\pm 1 \pm \tau^{-1}i \pm \tau j), & \tfrac{1}{2}(\pm 1 \pm \tau i \pm \tau^{-1}k), & \tfrac{1}{2}(\pm 1 \pm \tau^{-1}j \pm \tau k), \\ & & \tfrac{1}{2}(\pm i \pm \tau j \pm \tau^{-1}k), \\ \tfrac{1}{2}(\pm \tau \pm i \pm \tau^{-1}j), & \tfrac{1}{2}(\pm \tau \pm \tau^{-1}i \pm k), & \tfrac{1}{2}(\pm \tau \pm j \pm \tau^{-1}k), \\ & & \tfrac{1}{2}(\pm \tau i \pm \tau^{-1}j \pm k), \\ \tfrac{1}{2}(\pm \tau^{-1} \pm \tau i \pm j), & \tfrac{1}{2}(\pm \tau^{-1} \pm i \pm \tau k), & \tfrac{1}{2}(\pm \tau^{-1} \pm \tau j \pm k), \\ & & \tfrac{1}{2}(\pm \tau^{-1}i \pm j \pm \tau k). \end{cases}$$

These last quaternions are obtained by taking a regular dodecahedron of which one of the five inscribed cubes has vertices $(\pm 3^{-\frac{1}{2}}, \pm 3^{-\frac{1}{2}}, \pm 3^{-\frac{1}{2}})$. Thus $\frac{1}{2}(1 + \tau i + \tau^{-1}k)$ is obtained by writing $\xi = 3^{-\frac{1}{2}}\tau$, $\eta = 0$, $\zeta = 3^{-\frac{1}{2}}\tau^{-1}$, $\theta = \frac{1}{3}\pi$. Moreover, since every quaternion of unit norm represents a rotation, *every* finite group of quaternions (except the cyclic group generated by a single quaternion) can be transformed into one of the above groups.

When the quaternions are interpreted as points in four-dimensional space, (6.1) gives the vertices of two regular $2n$-gons in absolutely perpendicular planes (which, when $n = 2$, are also the vertices of a 16-cell $\{3, 3, 4\}$), (6.2) gives the vertices of a 24-cell $\{3, 4, 3\}$, (6.3) gives those of the reciprocal $\{3, 4, 3\}$, and (6.4) gives those of the *snub 24-cell*[9] $s\{3, 4, 3\}$, so that the 120 elements of the group $\langle 2, 3, 5\rangle$ correspond to the 120 vertices of the 600-cell[10] $\{3, 3, 5\}$, as has already been pointed out by Threlfall (loc. cit.).

[8] D. E. Littlewood, *The groups of the regular solids in n dimensions*, Proceedings of the London Mathematical Society, (2), vol. 32(1931), pp. 10–20; p. 13.

[9] H. S. M. Coxeter, *Wythoff's construction for uniform polytopes*, Proceedings of the London Mathematical Society, (2), vol. 38(1935), pp. 327–339; p. 338.

[10] P. H. Schoute, *Mehrdimensionale Geometrie*, vol. 2, Leipzig, 1905, p. 211.

Each of these sets of points in four dimensions has the property of being invariant under the reflection which interchanges any pair of opposite points. This is a consequence of the fact that the reflection which interchanges the pair of points ± 1 replaces each quaternion Q by $-Q^{-1}$. Thus a binary polyhedral group of order $2g$ corresponds to a symmetrical arrangement of $2g$ points on the unit hypersphere, and so to a set of g hyperplanes, reflections in which generate a group of orthogonal transformations in four dimensions. In detail, the results are as follows.

Binary polyhedral group	Order	Polytopes[11]	Group generated by reflections	Order
$\langle 2, 2, n\rangle$	$4n$	$\{2n\} + \{2n\}$	$[n] \times [n]$	$4n^2$
$\langle 2, 2, 2\rangle$	8	$\{3, 3, 4\}$	$[\ ] \times [\ ] \times [\ ] \times [\ ]$	16
$\langle 2, 3, 3\rangle$	24	$\{3, 4, 3\}$	$[3^{1,1,1}]$	192
$\langle 2, 3, 4\rangle$	48	$\{3, 4, 3\} + \{3, 4, 3\}$	$[3, 4, 3]$	1152
$\langle 2, 3, 5\rangle$	120	$\{3, 3, 5\}$	$[3, 3, 5]$	14400

The dicyclic group[12] $\langle 2, 2, n\rangle$, in the form

$$S^2 = T^n = (ST)^2,$$

can be derived from the cyclic group $T^{2n} = 1$ by adjoining an element S such that $S^{-1}TS = T^{-1}$, $S^2 = T^n$. The binary polyhedral groups $\langle 2, 3, n\rangle$, for $n = 3, 4, 5$, can be represented by permutations of degrees 8, 16, 24, corresponding to the representation of the ordinary polyhedral groups $(2, 3, n)$ by permutations of the vertices of the tetrahedron, cube, and icosahedron.

7. **The search for a subgroup** $\langle | l |, | m |, | n |\rangle$. Consider the general group $\langle l, m, n\rangle$, defined by (1.2) with

$$| l | > 1, \quad | m | > 1, \quad | n | > 1, \quad | l |^{-1} + | m |^{-1} + | n |^{-1} > 1.$$

Let us see whether it is possible to find integers r, s, t such that

$$R = R_0Z^r, \quad S = S_0Z^s, \quad T = T_0Z^t, \tag{7.1}$$

where

$$R_0^{|l|} = S_0^{|m|} = T_0^{|n|} = R_0S_0T_0 .$$

Since Z is permutable with R, S, T, we have

$$R_0^{|l|} = R^{|l|}Z^{-|l|r} = Z^{|l|(l-1-r)}$$

[11] H. S. M. Coxeter, *Finite groups generated by reflections, and their subgroups generated by reflections*, Proceedings of the Cambridge Philosophical Society, vol. 30(1934), pp. 466–482; p. 468.

[12] G. A. Miller, H. F. Blichfeldt, and L. E. Dickson, *Finite Groups*, New York, 1916, p. 62.

and similarly

$$S_0^{|m|} = Z^{|m|(m^{-1}-s)}, \quad T_0^{|n|} = Z^{|n|(n^{-1}-t)}, \quad R_0S_0T_0 = Z^{1-r-s-t}.$$

Equating these four powers of Z (after a change of sign), we get

$$|l|(r - l^{-1}) = |m|(s - m^{-1}) = |n|(t - n^{-1}) = r + s + t - 1.$$

If each of these expressions is equal to u, we have

$$(|l|^{-1} + |m|^{-1} + |n|^{-1} - 1)u = (r - l^{-1}) + (s - m^{-1}) + (t - n^{-1})$$
$$- (r + s + t - 1) = 1 - l^{-1} - m^{-1} - n^{-1},$$

and therefore

$$u = (1 - l^{-1} - m^{-1} - n^{-1})(|l|^{-1} + |m|^{-1} + |n|^{-1} - 1)^{-1}.$$

Table of Values of u

($m = 2$)

l \ n	-3	2	3	4	5
-5	31	1	11		
-4	13	1	5		
-3	7	1	**3**	7	19
-2	**4**	1	**2**	3	**4**

($m = -2$)

l \ n	-3	-2	2	3	4	5
-5	61	11	**6**	41		
-4	25	9	5	17		
-3	13	7	**4**	**9**	19	49
-2	7	5	3	5	7	9
3	**9**	5	**2**	5	11	29

The numbers

$$r = u|l|^{-1} + l^{-1}, \quad s = u|m|^{-1} + m^{-1}, \quad t = u|n|^{-1} + n^{-1}$$

are integers if and only if[13]

$$u + \operatorname{sgn} l, \qquad u + \operatorname{sgn} m, \qquad u + \operatorname{sgn} n$$

are divisible by $|l|$, $|m|$, $|n|$, respectively. On referring to the above table for u, we see that the cases of failure (marked in boldface type) are:

(i) $\langle -2, 2, n\rangle$, n odd, $\quad u = |n - 1|$;

(ii) $\langle -3, 2, 3\rangle$, $\quad u = 3$;

(iii) $\langle -3, -2, 3\rangle$, $\quad u = 9$.

Hence, in all save these exceptional cases, the group $\langle l, m, n\rangle$ contains elements

$$R_0 = R^{-u \operatorname{sgn} l}, \quad S_0 = S^{-u \operatorname{sgn} m}, \quad T_0 = T^{-u \operatorname{sgn} n},$$

such that

(7.2) $$R_0^{|l|} = S_0^{|m|} = T_0^{|n|} = R_0S_0T_0 = Z^{-u}.$$

[13] We use the abbreviation $\operatorname{sgn} l = l/|l|$, i.e., ± 1 according to the sign of l.

It then follows from §4 that

$$Z^{2u} = 1, \tag{7.3}$$

i.e., that the period of Z is a divisor of $2u$.

It remains to be proved that (7.3) continues to hold in the three exceptional cases.

8. **The groups $\langle -2, 2, n\rangle$ and $\langle -3, \pm 2, 3\rangle$.**

(i) For $\langle -2, 2, n\rangle$, the defining relations

$$R^2 = S^2 = (RS)^n = Z \tag{8.1}$$

imply $RS \cdot SR = RZR = R^2Z = Z^2$, whence

$$(RS)^n(SR)^n = Z^{2n}.$$

But $(RS)^n = Z = (SR)^n$. Therefore

$$Z^{2n-2} = 1,$$

and the period of Z is a divisor of $| 2n - 2 |$.

(ii) For $\langle -3, 2, 3\rangle$, the defining relations

$$R^2 = S^3 = (RS)^3 = Z$$

imply[14] $RS^{-1}R^{-1}S = ZR^{-1}S^{-1}R^{-1}S = SRS^2$, whence

$$(RS^{-1}R^{-1}S)^2 = SRS^3RS^2 = Z^3,$$

and

$$Z^6 = Z^3 \cdot R^{-1}Z^3R = (R^{-1}S^{-1}RS)^2(S^{-1}R^{-1}SR)^2 = 1.$$

(iii) Putting S^{-1}, T^{-1} for S, T in (1.3), we see that $\langle -3, -2, 3\rangle$ is defined by

$$S^3 = T^2 = Z, \qquad (ST)^3 = Z^{-2},$$

whence

$$TS^{-1}T^{-1}S = ZT^{-1}S^{-1}T^{-1}S = Z^3STS^2 = Z^4STS^{-1},$$

$$(TS^{-1}T^{-1}S)^2 = Z^8ST^2S^{-1} = Z^9,$$

and

$$Z^{18} = Z^9 \cdot T^{-1}Z^9T = (T^{-1}S^{-1}TS)^2(S^{-1}T^{-1}ST)^2 = 1.$$

9. **Proof that the period of Z is not less than $2u$.** The element Z^2 of $\langle l, m, n\rangle$ generates an invariant cyclic subgroup, whose quotient group,

$$R^l = S^m = T^n = RST = Z, \qquad Z^2 = 1,$$

is $\langle | l |, | m |, | n |\rangle$, by §5. Therefore the period of Z is *even*.

[14] This argument was suggested by Miller, Blichfeldt, and Dickson, op. cit., p. 153.

Another factor group of $\langle l, m, n\rangle$ (the "commutator quotient group") is obtained from (1.3) by inserting the relation

$$ST = TS,$$

so as to make the group Abelian. This factor group is cyclic whenever two of l, m, n, say l and m, are coprime. For, we can then find integers λ and μ, likewise coprime, such that

$$\lambda m + \mu l = 1.$$

Since $S^l = Z^{l-1}T^{-l}$, this implies

$$S = S^{\lambda m+\mu l} = Z^{\lambda+\mu(l-1)}T^{-\mu l} = T^{\lambda n+\mu(nl-n-l)}$$

and

$$S^m = T^{(1-\mu l)n+\mu m(nl-n-l)} = T^{n+\mu(lmn-mn-nl-lm)}.$$

Similarly

$$R = T^{\mu n+\lambda(mn-m-n)} \qquad \text{and} \qquad R^l = T^{n+\lambda(lmn-mn-nl-lm)}.$$

Hence the cyclic group generated by T is of order $|\, lmn - mn - nl - lm\,|$.

Let c be the least common multiple of $|\, l\,|$, $|\, m\,|$, $|\, n\,|$, and write

$$h = 1 - l^{-1} - m^{-1} - n^{-1}.$$

Since l and m are coprime, c is the least common multiple of $|\, lm\,|$ and $|\, n\,|$. Since the period of T is $|\, lmnh\,|$, that of Z $(= T^n)$ is the least integral multiple of $|\, lmh\,|$. Since lmh involves the fraction lm/n, the required multiplier is $|\, n\,|/(|\, lm\,|, |\, n\,|) = c/|\, lm\,|$, and the period of Z is ch.

Among the groups we are considering, the only cases in which no two of l, m, n are coprime are those in which $|\, l\,| = |\, m\,| = 2$, while n is even. These are covered by the supposition that l divides m, in which case we consider (instead of the commutator quotient group) the cyclic factor group given by the extra relation $R = S^{m/l}$. This reduces (1.1) to

$$S^m = T^n = S^{(m/l)+1}T,$$

whence

$$T = S^{m-(m/l)-1} \qquad \text{and} \qquad S^{mn-(mn/l)-n-m} = 1.$$

This factor group is thus the cyclic group of order $|\, mnh\,|$, and the period of Z is the least integral multiple of $|\, nh\,|$. Since nh involves the fraction n/m, the required multiplier is $c/|\, n\,|$ (where c is the least common multiple of $|\, m\,|$ and $|\, n\,|$), and the period of Z is again ch.

Returning to the whole group $\langle l, m, n\rangle$, we conclude that the period of Z is divisible by ch. Moreover, since ch is odd in every case except when $|\, l\,| = |\, m\,| = 2$ while n is odd, we may put together our two results by saying that

the period of Z is divisible by $2ch$, except in that special case. We observe also that

$$c^{-1} = |\,l\,|^{-1} + |\,m\,|^{-1} + |\,n\,|^{-1} - 1,$$

except in the same special case. Hence, in *every* case, the period of Z is divisible by $2h(|\,l\,|^{-1} + |\,m\,|^{-1} + |\,n\,|^{-1} - 1)^{-1}$.

10. **General formula for the order of $\langle l, m, n\rangle$.** From §§7, 8, 9, we conclude that, so long as $|\,l\,|$, $|\,m\,|$, $|\,n\,|$, and $|\,l\,|^{-1} + |\,m\,|^{-1} + |\,n\,|^{-1}$ are all greater than 1, the period of Z in $\langle l, m, n\rangle$ is exactly

$$2u = |\,l^{-1} + m^{-1} + n^{-1} - 1\,|\,g,$$

where

$$g = 2(|\,l\,|^{-1} + |\,m\,|^{-1} + |\,n\,|^{-1} - 1)^{-1};$$

and consequently *the order of $\langle l, m, n\rangle$ is*

$$|\,l^{-1} + m^{-1} + n^{-1} - 1\,|\,g^2.$$

11. **Direct products.** We proceed to prove that, apart from the exceptional cases $\langle -2, 2, n\rangle$ (n odd) and $\langle -3, \pm 2, 3\rangle$, each group $\langle l, m, n\rangle$ (with $|\,l\,|$, $|\,m\,|$, $|\,n\,|$, and $|\,l\,|^{-1} + |\,m\,|^{-1} + |\,n\,|^{-1}$ all greater than 1) is *the direct product of $\langle |\,l\,|, |\,m\,|, |\,n\,|\rangle$ and the cyclic group of order u.*

Since u is odd, and Z^u is given by (7.2), we see at once that $\langle l, m, n\rangle$ is generated by R_0, S_0, T_0, and Z^2. Moreover, R_0, S_0, T_0 generate a group whose order divides $2g$, and Z^2 generates the cyclic group of order u. But the order of $\langle l, m, n\rangle$ is $2gu$. Hence R_0, S_0, T_0 generate $\langle |\,l\,|, |\,m\,|, |\,n\,|\rangle$, and $\langle l, m, n\rangle$ is the direct product of this and the cyclic group generated by Z^2.

12. **Further remarks on $\langle -2, 2, n\rangle$ and $\langle -3, 2, 3\rangle$.** It follows from §10 that the group $\langle -2, 2, n\rangle$, defined by (8.1), is of order $4n(n-1)$, while its element RS is of period $2n(n-1)$. Thus the cyclic subgroup generated by RS is of index two. Moreover, since $SR = (RS)^{-1}Z^2 = (RS)^{2n-1}$, the whole group can be derived from this subgroup by adjoining an element R, whose square is $(RS)^n$, and which transforms each element of the subgroup into its $(2n-1)$-th power.

If n is even, the subgroup is the direct product of the cyclic groups generated by

$$(RS)^{2n} = Z^2 \qquad \text{and} \qquad (RS)^{n-1} = ZS^{-1}R^{-1} = SR^{-1} = S^{-1}R.$$

The extra element R is permutable with the former cyclic group, but transforms each element of the latter into its inverse. This continues to hold if we replace R by R^{n-1}, whose square is $Z^{n-1} = (S^{-1}R)^n$. The whole group is therefore the direct product of the cyclic group generated by Z^2 and the dicyclic group $\langle 2, 2, n\rangle$, as we saw in §11.

If, on the other hand, n is odd, the subgroup generated by RS is the direct product of the cyclic groups generated by

$$(RS)^n = Z \quad \text{and} \quad (RS)^{2(n-1)} = (SR^{-1})^2 = (S^{-1}R)^2.$$

The extra element R is permutable with the former cyclic group, and transforms each element of the latter into its inverse.[15] But since R^2 $(= Z)$ belongs to the former cyclic group and not to the latter, this no longer makes $\langle -2, 2, n\rangle$ a direct product. Perhaps the simplest description is as follows. When n is odd, $\langle -2, 2, n\rangle$ is derived from the cyclic group of order n by adjoining an element R, of period $4(n - 1)$, which transforms every element of the cyclic group into its inverse.

A. Sinkov has obtained the following information about $\langle -3, 2, 3\rangle$ by representing it as a regular permutation group. It has four Sylow subgroups of order 9, and one of order 8. Miller[16] has shown that there are four such groups of order 72; $\langle -3, 2, 3\rangle$ is that one of the four in which the subgroups of order 9 are cyclic while the subgroup of order 8 is the quaternion group. Moreover, since there is only one subgroup of order 3, and one of order 2, both these are invariant, and at least one of them is contained in any subgroup that may be formed. Hence no representation is possible on fewer than 72 letters.

13. **Finite factor groups of $\langle 2, 3, 6\rangle$, $\langle 2, 4, 4\rangle$, $\langle 3, 3, 3\rangle$.** When $l^{-1} + m^{-1} + n^{-1} = 1$, the group $\langle l, m, n\rangle$ is, of course, infinite. By generalizing the method used by Burnside[17] for the corresponding groups (l, m, n), J. M. Kingston has obtained the following finite factor groups of $\langle 2, 3, 6\rangle$, $\langle 2, 4, 4\rangle$, $\langle 3, 3, 3\rangle$:

$$S^3 = T^6 = (ST)^2 = Z, \quad (ST^{-2})^b(S^{-1}T^2)^c = Z^a = 1,$$

of order $6a(b^2 + bc + c^2)$;

$$S^4 = T^4 = (ST)^2 = Z, \quad (ST^{-1})^b(S^{-1}T)^c = Z^a = 1,$$

of order $4a(b^2 + c^2)$;

$$S^3 = T^3 = Z, \quad (ST)^3 = Z^2, \quad (ST^{-1})^b(S^{-1}T)^c = Z^a = 1,$$

of order $3a(b^2 + bc + c^2)$.

[15] Cf. G. A. Miller, *The groups generated by two operators which have a common square*, Archiv der Mathematik und Physik, (3), vol. 9(1905), pp. 6–7.

[16] G. A. Miller, *Determination of all the abstract groups of order* 72, American Journal of Mathematics, vol. 51(1929), pp. 491–494, first paragraph of section III.

[17] W. Burnside, *Theory of Groups of Finite Order*, Cambridge, 1911, pp. 413–419.

In the cases $b = 0$ and $b = c$, he has represented these three groups by permutations, of the following degrees:

$$(b = 0) \begin{cases} 6\,ac, \\ 4\,ac, \\ 3\,ac, \end{cases} \qquad (b = c) \begin{cases} 18\,ac, \\ 8\,ac, \\ 9\,ac. \end{cases}$$

Corresponding results when $a = 1$ were obtained by Edington,[18] Sinkov,[19] and Coxeter.[20]

UNIVERSITY OF TORONTO.

[18] W. E. Edington, *Abstract group definitions and applications*, Transactions of the American Mathematical Society, vol. 25(1923), pp. 193–210.

[19] A. Sinkov, *Notes on the groups of genus one*, Tôhoku Mathematical Journal, vol. 43(1937), pp. 164–170.

[20] H. S. M. Coxeter, *The abstract groups $G^{m,n,p}$*, Transactions of the American Mathematical Society, vol. 45(1939), pp. 73–150; pp. 81, 83, 98–100.

TRILINEAR FORMS

By A. B. Coble

1. **Introduction.** We study the correspondence set up by the trilinear form

$$T(m, n, p) = (\alpha x)(\beta y)(\gamma z) = 0,$$

where x, y, z are points in spaces $[m]$, $[n]$, $[p]$ respectively with respective prime coördinates ξ, η, ζ. Thus

$$T = \sum_{i,j,k} \alpha_i \beta_j \gamma_k x_i y_j z_k = \sum_{i,j,k} a_{ijk} x_i y_j z_k$$

$$(i = 0, \cdots, m; j = 0, \cdots, n; k = 0, \cdots, p).$$

We are interested only in the projective properties of T, those invariant under *digredient* linear transformation of x, y, z. The cases in which two of the points lie in the same projective space, or in two different spaces which are projectively related, require special treatment. We assume initially that the variables are so named that

$$(1) \qquad 1 \leqq m \leqq n \leqq p,$$

though, occasionally, when only subspaces of $[n]$ or $[p]$ are under consideration, this convention must be modified.

We also assume initially that the correspondence $T = 0$ has no neutral points x, or y, or z, i.e., points for which the bilinear form in the remaining two variables vanishes identically. If, for example, T had neutral points x, these evidently would fill up a linear space $[k]$. If x were linearly transformed so that $k + 1$ of the new reference points are found in $[k]$, then $T(m, n, p)$ would be converted into a $T'(m - k - 1, n, p)$, and this simpler correspondence $T' = 0$ would be the proper subject for study. If for any purpose it were desirable to interpret a figure F' defined by T' in the space $[x'] = [m - k - 1]$ as a figure F in the original space $[x] = [m]$, it would be necessary only to *dilate* the figure F' in $[m - k - 1]$ into a figure F in $[m]$ by using the vertex $[k]$. Occasionally this procedure is necessary. The assumption that T has no neutral points z yields an upper limit for p when m, n are given, namely,

$$(2) \qquad p + 1 \leqq (m + 1)(n + 1).$$

2. **Neutral pairs of** T. We say that x, y is a pair of points neutral for z in the correspondence $T = 0$, if $(\alpha x)(\beta y)\gamma_k = 0$ for $k = 0, 1, \cdots, p$. With similar

Received July 9, 1940.

definitions for neutral pairs x, z and y, z we have to do with three systems of equations,

$$(1) \quad \text{(a)}\ \alpha_i(\beta y)(\gamma z) = 0; \qquad \text{(b)}\ (\alpha x)(\gamma z)\beta_j = 0; \qquad \text{(c)}\ (\alpha x)(\beta y)\gamma_k = 0,$$

with i, j, k defined as above. Since, because of the inequalities (1) of §1, $n + p \geqq m + 1$ and $m + p \geqq n + 1$, there always exist neutral pairs y, z and neutral pairs x, z. However, unless $m + n \geqq p + 1$, neutral pairs x, y will not necessarily exist.

In seeking to determine these neutral pairs we have obviously to deal with three matrices,

$$(2) \qquad \begin{aligned} &\text{(a)}\ M_{n,p}(x) = \mid (\alpha x)\beta_j\gamma_k \mid, \\ &\text{(b)}\ M_{m,p}(y) = \mid (\beta y)\alpha_i\gamma_k \mid, \\ &\text{(c)}\ M_{m,n}(z) = \mid (\gamma z)\alpha_i\beta_j \mid. \end{aligned}$$

It is clear that each one of these uniquely determines the form T to within projectivities, the operation of elementary transformations on the rows or on the columns corresponding to the operation of linear transformation on the one or the other of the two remaining variables. Thus to within unessential modifications each matrix uniquely determines the other two.

Each of the three systems of equations (1) defines two systems of linear spaces. For example, the $m + 1$ equations (1) (a) define, for given y, a linear subspace $[p - m - 1]$ of the space $[p]$ of z, which we denote by ${}_z[p - m - 1]_y$. Similarly, for given z, they define a subspace ${}_y[n - m - 1]_z$ of the space $[n]$ of y. Similarly, the $n + 1$ equations (1) (b) define two systems of subspaces ${}_z[p - n - 1]_x$, ${}_x[m - n - 1]_z$; and the $p + 1$ equations (1) (c) define ${}_y[n - p - 1]_x$, ${}_x[m - p - 1]_y$. It is to be emphasized that the dimensions of these subspaces are given for generic position of the x, or y, or z which defines them and that a negative value merely means non-existence for such generic position.

The points x, y, or z, for which the dimensions of the corresponding subspaces increase, are those for which the ranks of the matrices (2) (a), (b), (c) fall below their respective maxima, $n + 1$, $m + 1$, $m + 1$. We indicate the rank in question by a prefix attached to the matrix. Thus the locus of points x for which ${}_nM_{n,p}(x) = 0$, i.e., for which all the determinants of order $n + 1$ in $M_{n,p}(x)$ vanish, has the dimension $m + n - p - 1$ in $[m]$. This is also the dimension of ${}_mM_{m,p}(y) = 0$ in $[n]$. However, the dimension of ${}_mM_{m,n}(z) = 0$ in $[p]$ is $p + m - n - 1$. With reference to the inequalities (1) of §1 it is clear that this last dimension is zero only in the case $m = n = p = 1$, and in all other cases is positive. There will therefore always exist in $[p]$ a manifold ${}_mM_{m,n}(z) = 0$ of dimension $p + m - n - 1 < p$. This manifold is exhaustively studied in

Room's recent treatise[1] under the notation $V(|\, m+1, n+1\,|, p)$, the order of V being $\binom{n+1}{m}$. For a generic point z on ${}_mM_{m,n}(z) = 0$, a point for which ${}_{m-1}M_{m,n}(z) \neq 0$, there is a unique point x which satisfies (1) (b).

If p is greater than is implied by the limiting equality

$$(3) \qquad p + 1 = m + n,$$

the manifolds ${}_nM_{n,p}(x) = 0$ and ${}_mM_{m,p}(y) = 0$ will not exist, unless the form $T(m, n, p)$ is subject to such specializing conditions as ensure the existence. Let $x^{(0)}$ be a point for which $M_{n,p}(x)$ has the rank $n + 1 - k$ $(k > 0)$. Then (1) (c) has k independent solutions $y^{(0)}, \cdots, y^{(k-1)}$, i.e., there exists a space ${}_y[k-1]_{x^{(0)}}$. If $y^{(0)}$ is a particular point in ${}_y[k-1]_{x^{(0)}}$, then the subspace of $[m]$ defined by $y = y^{(0)}$ in (1) (c) exists and contains independent points $x^{(0)}, \cdots, x^{(l-1)}$ if $M_{m,p}(y)$ has the rank $m + 1 - l$ for $y = y^{(0)}$. This situation occurs, for example, when $x^{(0)}, y^{(0)} = e_0, f_0$, the first reference points in $[m]$, $[n]$ respectively, provided that $\alpha_0\beta_0(\gamma z) \equiv 0$ together with $\alpha_0\beta_1(\gamma z) \equiv \cdots \equiv \alpha_0\beta_{k-1}(\gamma z) \equiv 0$ and $\alpha_1\beta_0(\gamma z) \equiv \cdots \equiv \alpha_{l-1}\beta_0(\gamma z) \equiv 0$. Under these circumstances the equations (1) (a) for $y = y^{(0)}$ satisfy l linear relations with coefficients arising from $x^{(0)}, \cdots, x^{(l-1)}$, so that the space ${}_z[p-m-1]_y$ becomes for $y = y^{(0)}$ a space ${}_z[p-m-1+l]_{y^{(0)}}$, and similarly the space ${}_z[p-n-1]_x$ becomes for $x = x^{(0)}$ a space ${}_z[p-n-1+k]_{x^{(0)}}$. Like considerations apply to other subspaces. However, the normal situation for which the manifolds ${}_nM_{n,p} = 0$, ${}_mM_{m,p} = 0$ exist is that the two manifolds are birationally related by equations (1) (c), i.e., $k = l = 1$.

Particular cases in which some of the subspaces defined above are points are of interest. If the ${}_z[p-m-1]_y$ are points, so that $p = m + 1$, then two cases arise. Either T is a $T(m, m+1, m+1)$ and yields the Cremona transformation from $[y]$ to $[z]$ determined by $m + 1$ bilinear forms, or T is a $T(m, m, m+1)$ and yields a mapping of $[y]$ (or $[x]$) upon a determinantal primal of order $m + 1$ in $[z]$ (cf., for $m = 3$, Room, loc. cit., Chapter XV). If the ${}_z[p-n-1]_x$ are points and $p = n + 1$, then T is a $T(m, n, n+1)$, and we have a mapping of $[x]$ on an F_m in $[z] = [n+1]$. If the ${}_y[n-m-1]_z$ are points and $n = m + 1$, then T is a $T(m, m+1, m+1+k)$, where $k \geqq 0$, and we have a mapping of $[y]$ upon ∞^{m+1} spaces $[k]$ of $[z]$ which cover the space $[z]$ simply.

We shall refer to Room's determinantal locus of dimension $p + m - n - 1$ and order $\binom{n+1}{m}$ in the space $[p]$ as $V_{m,n,p}$. It is ordinarily the locus of its ∞^m generator spaces ${}_z[p-n-1]_x$. If, however, $p = n$, these generators do not exist for generic x. Then $V_{m,n,p}$ is the locus of its ∞^{m-1} generator points ${}_z[0]_{x^{(0)}}$, $x^{(0)}$ being a generic point on the manifold ${}_nM_{n,p}(x) = 0$. The pecu-

[1] T. G. Room, *The Geometry of Determinantal Loci*, Cambridge University Press, 1938.

liarity of the spaces ${}_z[p - m - 1]_y$ with respect to $V_{m,n,p}$ is expressed by the theorem:

(4) *A generic* $[p - m - 1]$ *in* $[p]$ *meets* $V_{m,n,p}$ *in a* $V_{m,n,p-m-1}$ *of dimension* $p - n - 2$ *and order* $\binom{n+1}{m}$; *a* ${}_z[p - m - 1]_y$ *meets* $V_{m,n,p}$ *in a* $V_{m,n-1,p-m-1}$ *of dimension* $p - n - 1$ *and order* $\binom{n}{m}$.

The first of these statements is obtained by restricting z in $[p]$ to the given $[p - m - 1]$; the second from the additional observation that, for z so restricted, y is a neutral point of the related $T(m, n, p - m - 1)$, and thus the space $[n]$ is effectively only a space $[n - 1]$. Room refers to the spaces ${}_z[p - m - 1]_y$ as the *second system of generators of* $V_{m,n,p}$.

3. Trilinear forms for which $p + 1 = m + n$. This case is of particular interest in that there is in general a finite number of neutral pairs x, y, this number being

$$N = \binom{m+n}{m} = \binom{m+n}{n}. \tag{1}$$

The points x make up a set P_N^m in $[m]$, and the points y a set Q_N^n in $[n]$, the two sets being ordered with respect to each other. This set P is the manifold ${}_nM_{n,p}(x) = 0$ and the set Q the manifold ${}_mM_{m,p}(y) = 0$.

We exclude the case $m = 1$ for which $p = n$. T is then merely a pencil (parameter x) of bilinear forms in y and z. P_N^1 is then the $n + 1$ points x for which the bilinear form has the rank n and neutral points y, z. Q_N^n is the set of $n + 1$ neutral points y, and $V_{m,n,p}$ is the set of $n + 1$ neutral points z.

With $m \geqq 2$, $n \geqq m$, and $p = m + n - 1$, the generator spaces ${}_z[p - n - 1]_x$ of $V_{m,n,p}$ with dimension $2m - 2$ and order $\binom{n+1}{m}$ exist for generic x. Also the system of second generators of $V_{m,n,p}$, the system of spaces ${}_z[p - m - 1]_y$, exists for generic y. However, if x, y is a neutral pair p_h, q_h $(h = 1, \cdots, N)$ of T, the corresponding generators of V expand into spaces ${}_z[p - n]_{p_h} = \pi_h$ and ${}_z[p - m]_{q_h} = \kappa_h$. The space π_h lies on V. The ∞^{p-n} spaces $[p - n - 1]$ in π_h are the limiting positions of spaces ${}_z[p - n - 1]_x$ as x approaches p_h in $[m]$ along some one of the $\infty^{m-1} = \infty^{p-n}$ directions through it. The space κ_h cuts V in a $V_{m,n-1,n-1}$ (cf. §2, (4)) of dimension $m - 1$ and order $\binom{n}{m}$. The $\infty^{p-m} = \infty^{n-1}$ spaces $[p - m - 1]$ in κ_h are the limiting positions of spaces ${}_z[p - m - 1]_y$ as y approaches q_h in $[n]$ along some one of the ∞^{n-1} directions through it. If p_h, q_h and $p_{h'}$, $q_{h'}$ are two different neutral pairs of T, then $(\alpha p_h)(\beta q_{h'})(\gamma z) = 0$ is a prime ρ in $[p] = [m + n - 1]$, which contains π_h and $\kappa_{h'}$. But in $\rho = [m + n - 2]$ these two spaces π_h and $\kappa_{h'}$ of dimension $m - 1$ and $n - 1$ respec-

tively must meet. The spaces π_h, $\kappa_{h'}$ with the incidences just mentioned make up the "double-N configuration" treated by Room (loc. cit., Chapter IV).

When $p + 1 = m + n$, the trilinear form $T(m, n, p)$ has $(m + 1)(n + 1) \cdot (m + n) - 1 - m(m + 2) - n(n + 2) - (m + n - 1)(m + n + 1) = (m + n)(mn - 1) - (m^2 + n^2)$ absolute constants. The set of points P_N^m in $[m]$ has $\left[\binom{m+n}{m} - (m+2)\right]m$ absolute constants. When $m = 2$, the number of absolute constants in both T and P is the same, namely, $n^2 + 3n - 6$. For larger values of m, the number of absolute constants in a generic set P_N^m is greater than the number found in T, whence the set P_N^m determined by T has special properties. We turn therefore to the case $m = 2$, and seek to determine $T(2, n, n + 1)$ in terms of a generic ternary set of points P_N^2. The greater part of Room's treatment in this particular case concerns the mutual relations between the spaces $[x]$ and $[z]$. We shall here be more interested in the relations between the spaces $[x]$ and $[y]$, and the sets of points P_N^2 and Q_N^n in them, particularly in view of the application to be made in a later paper to the set of nodes of a ternary rational curve.

4. **The trilinear form** $T(2, n, n + 1)$, $n \geqq 2$. It is convenient henceforth to replace N by the more specific

$$(1) \qquad N_n = \binom{n+2}{2}.$$

The N_n neutral pairs p_k, q_k of T lie in the sets $P_{N_n}^2$, $Q_{N_n}^n$ in the spaces [2] and $[n]$ of $[x]$ and $[y]$ respectively. Both T and $P_{N_n}^2$, as observed above, have $n^2 + 3n - 6$ absolute constants. However, the set $Q_{N_n}^n$ in $[n]$, having no more absolute constants than T, must be subject to at least $(n - 2)[N_{n-1} - 3]$ projective conditions. We shall find both algebraic and geometric statements for these conditions.

The simplest geometric situation is that which arises from the $n + 1$ equations (1) (b) of §2. These, for given x, determine a unique point z, the space ${}_z[0]_x$. If ζ is a prime on this point z, the equation of z is given by the vanishing of the bordered determinant

$$(2) \qquad \begin{vmatrix} (\alpha x)\beta_j\gamma_k \\ \zeta_k \end{vmatrix} = 0.$$

When x is at p_k of the set $P_{N_n}^2$, the $n + 1$ equations in z are dependent with multipliers q_k, and (2) vanishes identically in ζ. Hence, for given ζ, (2) is the equation of a curve $Q^{n+1}(\zeta)$ of the linear system (∞^{n+1}) of curves of order $n + 1$ on the set $P_{N_n}^2$. Thus (2) is the equation of the map of the plane $[x]$ upon the points ${}_z[0]_x$ of a surface $F_2^{N_n-1}$ in $[z]$. The prime ζ cuts $F_2^{N_n-1}$ in a curve $Q^{N_n-1}(\zeta)$ which is the map of $Q^{n+1}(\zeta)$. For $n = 2, 3, 4, \cdots$ this surface F_2^3 is the cubic surface in [3], the Bordiga surface F_2^6 in [4], and the White surfaces in $[n + 1]$ (cf. Room, loc. cit., Chapter XIV).

In this mapping, the directions about the point p_h map into the points of a line π_h on $F_2^{N_{n-1}}$. The curve Q_h^n, the n-ic on all of the points of $P_{N_n}^2$ except p_h, maps into the curve of order N_{n-2} in which the space $\kappa_h = [n-1]$ meets $F_2^{N_{n-1}}$ (cf. §3). Thus the double-N_n configuration and its incidences are obvious. Before going further into the geometry in the space $[y]$, we give an algebraic definition of the set $Q_{N_n}^n$ in terms of the given set $P_{N_n}^2$.

5. **The sets $P_{N_n}^2$, $Q_{N_n}^n$ of neutral pairs p_h, q_h of $T(2, n, n-1)$.** As noted earlier (cf. §2, (2) (a), (b)) these neutral pairs are found as the zeros of the matrix equations:

$$P_{N_n}^2 : \mid (\alpha x)\beta_j\gamma_k \mid_n ; \qquad Q_{N_n}^n : \mid \alpha_i(\beta y)\gamma_k \mid_2 . \tag{1}$$

Let the points of $P_{N_n}^2$ be given by the equations $(p_h\xi) = 0$ $(h = 1, \cdots, N_n)$. Since only $\binom{n+1}{2}$ of the powers $(p_h\xi)^{n-1}$ are linearly independent, these powers are connected by $n+1$ independent linear relations of the form

$$\sum_h q_{jh}(p_h\xi)^{n-1} \equiv 0 \qquad (j = 0, 1, \cdots, n). \tag{2}$$

If these are multiplied respectively by $\eta_0, \eta_1, \cdots, \eta_n$ and added, we get the double identity:

$$\sum_h (q_h\eta)\cdot(p_h\xi)^{n-1} \equiv 0 \qquad (\text{in } \xi \text{ and } \eta). \tag{3}$$

We have thus defined, obviously only to within a projectivity, a set of points $Q_{N_n}^n$ in the space $[n]$ of $[y]$. By polarizing this identity to obtain

$$\sum_h (q_h\eta)\cdot(p_h\xi)\cdot(p_h\xi')^{n-1} \equiv 0 \qquad (\text{in } \eta, \xi, \xi'), \tag{4}$$

we find, by equating the coefficients of ξ' to zero, that only $N_n - N_{n-2} = 2n + 1$ of the products $(q_h\eta)$, $(p_h\xi)$ are linearly independent. Hence all N_n of these products are apolar to $n+2$ linearly independent bilinear forms $(\alpha x)(\beta y)\gamma_k = 0$ $(k = 0, \cdots, n+1)$. Thus the pairs $x, y = p_h, q_h$ are neutral for a trilinear form $T(2, n, n+1) = (\alpha x)(\beta y)(\gamma z)$, in which the variables z also are defined only to within a projectivity. Since also a projectivity in x applied to the points of the given $P_{N_n}^2$ will not affect the coefficients q_{jh} in (2), the variables x, y, z in T are digredient.

The form T thus defined by $P_{N_n}^2$ is generic. For, if it were subject to projective conditions, its set $P_{N_n}^2$ of neutral points p_h belonging to neutral pairs $x, y = p_h, q_h$ could not have $n^2 + 3n - 6$ absolute constants, whereas the set $P_{N_n}^2$ used above to construct T with this neutral set was generic. Hence we have

(5) *The generic trilinear form $T(2, n, n+1)$ with N_n pairs x, y neutral for z is projectively defined by the generic set $P_{N_n}^2$ of points x of these neutral pairs. The set $Q_{N_n}^n$ of points y is projectively defined in terms of $P_{N_n}^2$ by the identity* (3).

An immediate interpretation of (3) yields the following geometric statement of the $(n-2)(N_{n-1}-3)$ projective conditions on $Q^n_{N_n}$:

(6) *The set $Q^n_{N_n}$, subject to $(n-2)(N_{n-1}-3)$ projective conditions in $[n]$, has an associated set of points R_{N_n} in a space $[N_{n-1}-1]$ with the equivalent geometric peculiarity that the set R_{N_n} is on a Veronesean $V_2^{(n-1)^2}$ in $[N_{n-1}-1]$.*

Indeed, the plane of $[x]$ is mapped on the Veronesean by the totality of curves of order $n-1$ (cf. Room, loc. cit., p. 15) in such a way that $(x\xi)^{n-1}$ is mapped into a point r of V_2 in $[N_{n-1}-1]$. Thus the set $P^2_{N_n}$ is mapped into a generic set R_{N_n} on V_2 by setting $(r_h\theta) \equiv (p_h\xi)^{n-1}$. Then the identity (3) becomes bilinear in η, θ and expresses merely that the sets Q_{N_n} and R_{N_n} are associated.[2] We may note that V_2 admits a collineation g_8 and thus depends upon $N^2_{n-1}-9$ constants. It is then easy to verify that the condition that R_{N_n} lies on such a V_2 imposes $(n-2)(N_{n-1}-3)$ projective conditions on R_{N_n}, and therefore also on its associated set $Q^n_{N_n}$.

We saw that the N_n products $(q_h\eta)\cdot(p_h\xi)$ in (4) satisfy the N_{n-2} linear identities obtained from the coefficients of ξ', and that this was sufficient to define T. As an obvious extension of the idea of simple association as used above, we may state that

(7) *If two sets $P^2_{N_n}$ and $Q^n_{N_n}$ are N_{n-2}-tuply associated, and if $P^2_{N_n}$ is a generic set, then $P^2_{N_n}$, $Q^n_{N_n}$ are neutral sets for a trilinear form $T(2, n, n+1)$.*

This statement does not exhaust the peculiarity of the identity (4) in that the special character of the coefficients of ξ' with respect to those of ξ in the individual terms is not utilized. For this set $(p_h\xi')^{n-2} \equiv (s_h\varphi)$, thus mapping the set $P^2_{N_n}$ on a set S_{N_n} on a Veronesean $V_2^{(n-2)^2}$ in a space $[N_{n-2}-1]$. We then find, by equating the coefficients of η in (4) to zero, that only $N_n-(n+1)$ of the N_n products $(p_h\xi)\cdot(s_h\varphi)$ are linearly independent. Thus all of the products are apolar to $3N_{n-2}-N_{n-1} = n(n-2)$ independent bilinear forms $B_f = (\delta_f x)(\epsilon_f s)$. Conversely, if the pairs p_h, s_h satisfy such forms, the coefficients q can be recovered, and the form T again obtained, this leading to the identity (4) and the given form of the points $(s_h\varphi)$. Hence we have

(8) *If $n(n-2)$ bilinear forms $B_f(2, N_{n-2}-1) = (\delta_f x)(\epsilon_f s) = 0$ $(f = 1, \cdots, n(n-2))$ are satisfied by N_n pairs x, s, and if the N_n points x are generic, then for every x there is an s satisfying the forms. The plane of x is mapped by the equations $B_f = 0$ upon a Veronesean $V_2^{(n-2)^2}$ of points s in $[N_{n-2}-1]$.*

(9) *If a matrix of $n(n-2)$ rows and N_{n-2} columns, whose elements are ternary linear forms in x, has the rank $N_{n-2}-1$ for a generic set of points $P^2_{N_n}$, then it has this rank for every point x and is reducible by elementary transformations of the matrix, and linear transformation on x, to the matrix of coefficients of y in the forms*

$$(x_1y_2 - x_2y_1)y_0^{l_0}y_1^{l_1}y_2^{l_2}, \qquad (x_2y_0 - x_0y_2)y_0^{l_0}y_1^{l_1}y_2^{l_2}, \qquad (x_0y_1 - x_1y_0)y_0^{m_0}y_1^{m_1},$$

where $l_0 + l_1 + l_2 = m_0 + m_1 = n-3$.

[2] A. B. Coble, *Associated sets of points*, Trans. Amer. Math. Soc., vol. 24(1922), pp. 1–20.

This last theorem represents a form in which the conditions $(p_h\xi)\cdot(s_h\varphi) \equiv (p_h\xi)\cdot(p_h\xi')^{n-2}$ can be satisfied. The first significant case is $n = 3$, $N_n = 10$. The identically vanishing determinant can then be exhibited as a skew-symmetric determinant.

Another definition of the set $Q^n_{N_n}$ in terms of the set $P^2_{N_n}$ can be obtained from the identity (3). Let $Q^{n+1}(\zeta)$ and $Q^{n+1}(\zeta')$ (cf. §4) be two $(n + 1)$-ics on $P^2_{N_n}$ and on a residual set $R^2_{N_{n-1}}$, $N_n + N_{n-1}$ being $(n + 1)^2$. If $(\beta x)^{n-2}$, $(\beta' x)^{n-2}$ are arbitrary curves of order $n - 2$, we have, in

$$(\beta x)^{n-2}\cdot Q^{n+1}(\zeta) + (\beta' x)^{n-2}\cdot Q^{n+1}(\zeta'),$$

the linear system of dimension $2N_{n-2} - 1$ of curves of order $2n - 1$ on these two sets. There is therefore a single linear identity connecting the $(2n - 1)$-th powers of the points of the two sets, namely,

$$\sum_{h=1}^{h=N_n} (p_h\xi)^{2n-1} + \sum_{l=1}^{l=N_{n-1}} (r_l\xi)^{2n-1} \equiv 0, \tag{10}$$

this being a consequence of the numerical relation $N_{2n-1} = [(n + 1)^2 - 1] + 2N_{n-2}$.

Let $(\alpha x)^n(a\eta) = 0$ be the linear system, with $n + 1$ parameters η of n-ic curves on the set $R^2_{N_{n-1}}$. The polar of $(\alpha x)^n(a\eta)$ with respect to the identity (10), with x operating on ξ, yields the identity

$$\sum_{h=1}^{h=N_n} (\alpha p_h)(a\eta)\cdot(p_h\xi)^{n-1} \equiv 0. \tag{11}$$

On comparing this identity with (3) we have

$$(q_h\eta) \equiv (\alpha p_h)^n\cdot(a\eta). \tag{12}$$

An immediate interpretation of this result is as follows:

(13) *The set $Q^n_{N_n}$ is the map in $[n]$ of $P^2_{N_n}$ by n-ic curves on the residual base $R^2_{N_{n-1}}$ of a pencil of curves $Q^{n+1}(\zeta)$ on $P^2_{N_n}$. The plane of $[x]$ maps into a White surface $F_2^{N_{n-2}}$ and the pencil $Q^{n+1}(\zeta)$ maps into a pencil of curves $S^{N_{n-1}}(\zeta)$ on $F_2^{N_{n-2}}$, the base of this pencil S being $Q^n_{N_n}$.*

We proceed to examine the system of curves S in relation to the systems Q mentioned earlier.

6. The curves $Q^{n+1}(\zeta)$ on $P^2_{N_n}$, $Q^{N_{n-1}}(\zeta)$ on $F_2^{N_{n-1}}$, and $S^{N_{n-1}}(\zeta)$ on $Q^n_{N_n}$.

To every section $Q^{N_{n-1}}(\zeta)$ of the White surface $F_2^{N_{n-1}}$ in $[z]$ by a prime ζ, there corresponds a curve $Q^{n+1}(\zeta)$ of the linear system of planar $(n + 1)$-ics on $P^2_{N_n}$. The generic curve $Q^{n+1}(\zeta)$ of this system is a generic curve only so far as its order is concerned. It depends upon $N_{n+1} - 9$ absolute projective constants. It is, however, of high genus, $p = N_{n-2}$, with respect to its order and thus is birationally special if $n > 3$. Its peculiarity is that it contains a complete linear series, its line sections $L \equiv g_2^{n+1}$, such that its canonical series C is given by

$$C \equiv (n - 2)L. \tag{1}$$

The $N_{n+1} - 9$ absolute projective constants of $Q^{n+1}(\zeta)$ are the birational moduli both of $Q^{n+1}(\zeta)$ and of $Q^{N_n-1}(\zeta)$.

Let the prime ζ be fixed in the equation (2) of §4. Then the birationally related curves $Q^{N_n-1}(\zeta)$, $Q^{n+1}(\zeta)$ are also fixed. If $x^{(0)}$ is a particular point on $Q^{n+1}(\zeta)$ for which the determinant (2) of §4 vanishes, the bordered determinant must factor, i.e.,

$$(2)\qquad \begin{vmatrix} (\alpha x^{(0)})\beta_j\gamma_k & \eta_i \\ \zeta_k, \zeta'_k & 0 \end{vmatrix} = (z^{(0)}\zeta')\cdot(y^{(0)}\eta).$$

Indeed, this bordered determinant is the result of eliminating z and ρ from the relations

$$(3)\qquad (\alpha x^{(0)})(\beta y)(\gamma z) \equiv \rho(\eta y), \qquad (\zeta z) = 0, \qquad (\zeta' z) = 0,$$

the first relation being an identity in y. If ζ' is on the point $z^{(0)}$ of $Q^{N_n-1}(\zeta)$ which corresponds to $x^{(0)}$ on $Q^{n+1}(\zeta)$, the equations (3) are satisfied by $z = z^{(0)}$, $\rho = 0$, and this identifies the factor $(z^{(0)}\zeta')$. If $\rho \neq 0$, then, for every z on the $[n-1] = (\zeta, \zeta')$, the primes $(\alpha x^{(0)})(\beta y)(\gamma z) = 0$ in the space $[y]$ pass through a point $y^{(0)}$. As $x^{(0)}$ runs over $Q^{n+1}(\zeta)$, the points $z^{(0)}$, $y^{(0)}$ run over curves in $[z]$, $[y]$ respectively. For given ζ', η the points $z^{(0)}$, $y^{(0)}$ correspond to points $x^{(0)}$ cut out on $Q^{n+1}(\zeta)$ by a curve of order n. For fixed ζ', and therefore N_{n-1} fixed points on $Q^{N_n-1}(\zeta)$ and $Q^{n+1}(\zeta)$, and variable η, the locus of points $y^{(0)}$ is the map of $Q^{n+1}(\zeta)$ by a system of curves of order n, with parameters η, on the fixed N_{n-1} points and therefore also on a variable set of N_{n-1} points. Hence the locus of $y^{(0)}$ is a curve $S^{N_n-1}(\zeta)$ birationally related to $Q^{n+1}(\zeta)$.

If p_h, q_h is a neutral pair of T, and if η is on q_h, then (3) is satisfied for any ζ, ζ' since $(\alpha p_h)(\beta q_h)\gamma_k = 0$, $(\eta q_h) = 0$. Also if (ζ, ζ') meets at $z^{(0)}$ the line π_h on $F_2^{N_n-1}$ which corresponds to directions about p_h in $[x]$, then (3) is satisfied for any η with $\rho = 0$. Hence

$$(4)\qquad \begin{vmatrix} (\alpha p_h)\beta_j\gamma_k & \eta_i \\ \zeta_k, \zeta'_k & 0 \end{vmatrix} = (\pi_h\zeta\zeta')\cdot(q_h\eta).$$

Thus as $x^{(0)}$ on $Q^{n+1}(\zeta)$ passes through p_h, $z^{(0)}$ on $Q^{N_n-1}(\zeta)$ crosses π_h, and $y^{(0)}$ on $S^{N_n-1}(\zeta)$ passes through q_h. Hence

(5) *Each curve of the linear system* (∞^{n+1}) *of* $(n+1)$*-ics* $Q^{n+1}(\zeta)$ *on* $P^2_{N_n}$ *is birationally related to a curve* $Q^{N_n-1}(\zeta)$ *of the system* (∞^{n+1}) *of curves cut out on* $F_2^{N_n-1}$ *by primes* ζ *in* $[z]$, *and to a curve* $S^{N_n-1}(\zeta)$ *of a system* (∞^{n+1}) *on* $Q^n_{N_n}$ *in* $[y]$. *Prime sections of* $Q^{N_n-1}(\zeta)$ *correspond to sets* $R^2_{N_{n-1}}$ *on* $Q^{n+1}(\zeta)$ *which are cut out by* $(n+1)$*-ics on* $P^2_{N_n}$. *Prime sections of* $S^{N_n-1}(\zeta)$ *correspond to sets cut out on* $Q^{n+1}(\zeta)$ *by* n*-ics on a set* $R^2_{N_{n-1}}$.

We remark first that, if $n > 3$, the situation that $Q^n_{N_n}$ carries ∞^{n+1} curves $S^{N_n-1}(\zeta)$ implies conditions on the set of points $Q^n_{N_n}$. For, the curve $Q^{n+1}(\zeta)$ has $N_{n+1} - 9$ moduli, and the choice of a $g_n^{N_n-1}$ on it adds $p = N_{n-2}$ constants.

Thus the map $S^{N_{n-1}}(\zeta)$ has $N_{n+1} + N_{n-2} - 9$ absolute projective constants and a freedom of $N_{n+1} + N_{n-2} - 9 + n(n+2) = 2n^2 + 4n - 6$. However, when $n > 3$, this is less than the number $(n-1)N_n + n + 1$ of constants required to have a system ∞^{n+1} on a set of N_n generic points. For $n = 3$, these numbers coincide. Nevertheless for $n \geqq 3$ we have the special character of the set $Q^n_{N_n}$ observed in (13) of §5:

(6) *Any two curves of the system* (∞^{n+1}) *of curves* $S^{N_{n-1}}(\zeta)$ *on* $Q^n_{N_n}$ *lie in a pencil of such curves on a White surface* $F_2^{N_{n-2}}$.

We remark also that the two equations $T(2, n, n+1) = (\alpha x)(\beta y)(\gamma z) = 0$ and $(\zeta z) = 0$ are effectively the same as a single equation $T'(2, n, n) = 0$, where the $[n]$ of z in T' is ζ in the $[n+1]$ of T. For each x, $T' = 0$ is a correlation between the spaces $[y]$ and ζ, which degenerates when x is a point $x^{(0)}$ on $Q^{n+1}(\zeta)$, and then has singular points $y^{(0)}$, $z^{(0)}$ in $[y]$, ζ respectively. In this form the symmetry between the curves $Q^{N_{n-1}}(\zeta)$ and $S^{N_{n-1}}(\zeta)$ is more apparent.

The birational definition of the set $Q^n_{N_n}$ on a curve $S^{N_{n-1}}(\zeta)$ is a consequence of the planar equivalences which follow from (5):

$$P^2_{N_n} + R^2_{N_{n-1}} \equiv (n+1)L, \qquad \pi + R^2_{N_{n-1}} \equiv nL,$$

where π is a prime section of $S^{N_{n-1}}$. From these, the mapping (13) of §5 and (1), we have

$$Q^n_{N_n} \equiv \pi + L \equiv \pi + \frac{C}{n-2}. \tag{7}$$

Hence

(8) *If* $S^{N_{n-1}}(\zeta)$ *on* $Q^n_{N_n}$ *with prime section* π *is birationally related to* $Q^{n+1}(\zeta)$ *on* $P^2_{N_n}$ *with line section* L, *then the set* $Q^n_{N_n}$ *on* $S^{N_{n-1}}(\zeta)$ *has the birational definition* (7).

Thus the freedom of the set $Q^n_{N_n}$ on given S is $N_n - N_{n-2} = 2n + 1$. The freedom of S itself, as mentioned above, is $2n^2 + 4n - 6$, whence the freedom of incident Q_{N_n}, S is $2n^2 + 6n - 5$. On the other hand, the freedom of $Q^n_{N_n}$ in $[n]$ is $n^2 + 3n - 6 + n(n+2) = 2n^2 + 5n - 6$, and the freedom of S on $Q^n_{N_n}$ is $n + 1$, so that again the freedom of incident $Q^n_{N_n}$, S is $2n^2 + 6n - 5$.

7. **The N_{n-2}-secant $[n-2]$'s of the curves** $Q^{N_{n-1}}(\zeta)$, $S^{N_{n-1}}(\zeta)$. In our present case, $V_{m,n,p}$ is the White surface $F_2^{N_{n-1}}$ in $[z]$, and, according to (4) of §2, the ∞^n spaces ${}_z[n-2]_y$ obtained by fixing y in the equations (1) (a) of §2 cut $F_2^{N_{n-1}}$ in N_{n-2} points. They are the only spaces $[n-2]$ which cut $F_2^{N_{n-1}}$ in this manner. A particular ${}_z[n-2]_y$ is on three primes ζ, ζ', ζ'' of a net. Thus the N_{n-2} points z arise from a set of N_{n-2} points x, say $C^2_{N_{n-2}}(y)$, which with $P^2_{N_n}$ are the base of a net of curves $Q^{n+1}(\zeta)$. A particular curve $Q^{n+1}(\zeta)$ of this net is cut by the pencil $\lambda' Q^{n+1}(\zeta') + \lambda'' Q^{n+1}(\zeta'')$ on $P^2_{N_n}$ and $C^2_{N_{n-2}}(y)$ in a g^n_1 which is necessarily special, and therefore is on a pencil of lines through a point p of $Q^{n+1}(\zeta)$. Conversely, let g^n be any collinear n-point on a line L. This will be on a pencil of curves $Q^{n+1}(\zeta)$, say $\lambda Q^{n+1}(\zeta) + \lambda' Q^{n+1}(\zeta')$, which meet again

in a set $C^2_{N_n-2}$. Since g^n on $Q^{n+1}(\zeta)$ is in a complete g^n_1, there will be a pencil $\lambda' Q^{n+1}(\zeta') + \lambda'' Q^{n+1}(\zeta'')$ which cuts $Q^{n+1}(\zeta)$ in $P^2_{N_n}$, $C^2_{N_n-2}$, and the g^n_1. Since $C^2_{N_n-2}$ and the set g^n are a set $R^2_{N_n-1}$ of (5) of §6 on $Q^{n+1}(\zeta)$, and since ∞^{n-1} of the ∞^n n-ic curves on this set $R^2_{N_n-1}$ contain L, and therefore the residual point p, the $Q^{n+1}(\zeta)$ maps into a curve $S^{N_n-1}(\zeta)$ and the point p maps into the point y on $S^{N_n-1}(\zeta)$. Hence

(1) *The ∞^n spaces ${}_z[n-2]_y$ are the N_{n-2}-secant spaces of $F_2^{N_n-1}$ and of the curves $Q^{N_n-1}(\zeta)$ on $F_2^{N_n-1}$. Each ${}_z[n-2]_y$ is N_{n-2}-secant to the ∞^2 curves $Q^{N_n-1}(\zeta)$ which correspond to the ∞^2 curves $S^{N_n-1}(\zeta)$ on y, and to the ∞^2 curves $Q^{n+1}(\zeta)$ on $C^2_{N_n-2}(y)$. The ∞^1 N_{n-2}-secant $[n-2]$'s of a particular curve $Q^{N_n-1}(\zeta)$ are in one-to-one correspondence with the points y of $S^{N_n-1}(\zeta)$ and with the ∞^1 sets $C^2_{N_n-2}(y)$ on $Q^{n+1}(\zeta)$ which carry a net of such curves.*

If we fix z in the three equations (1) (a) of §2, $\alpha_i(\beta y)(\gamma z) = 0$, there is determined in $[y]$ the space ${}_y[n-3]_z$. If, however, $z^{(0)}$ is a point on $F_2^{N_n-1}$ which arises from $x^{(0)}$ in [2], so that $(\alpha x^{(0)})(\gamma z^{(0)})\beta_j = 0$, then the three equations above are linearly dependent in y with multipliers $x^{(0)}$, and thus we have a space ${}_y[n-2]_{z^{(0)}}$. The ∞^2 spaces of this character will be called Semple $[n-2]$'s [cf. Room, loc. cit., 14.7, p. 381, for the case $n = 3$]. A particular Semple $[n-2]$ is determined by either the $x^{(0)}$ in [2], or the $z^{(0)}$ in $[n+1]$ on $F_2^{N_n-1}$. Let ζ be a particular prime in $[n+1]$ on $z^{(0)}$, determining a curve $Q^{N_n-1}(\zeta)$ on $F_2^{N_n-1}$ through $z^{(0)}$, and a curve $Q^{n+1}(\zeta)$ on $P^2_{N_n}$ through $x^{(0)}$. Then, as remarked after (6) of §6, the equations $T = (\alpha x)(\beta y)(\gamma z) = 0$, $(\zeta z) = 0$ are effectively those of a $T'(2, n, n)$ in $[x]$, $[y]$, ζ. For $x^{(0)}$ on $Q^{n+1}(\zeta)$, the correlation T' between y and z is singular with singular points $y^{(0)}$, $z^{(0)}$ on $S^{N_n-1}(\zeta)$, $Q^{N_n-1}(\zeta)$ respectively. Just as, according to (1), for $y = y^{(0)}$ there is a space ${}_z[n-2]_{y^{(0)}}$ which is N_{n-2}-secant to $Q^{N_n-1}(\zeta)$, so, for $z = z^{(0)}$, there is a space ${}_y[n-2]_{z^{(0)}}$ which is N_{n-2}-secant to $S^{N_n-1}(\zeta)$, and this evidently is the Semple $[n-2]$ determined by $z^{(0)}$ or $x^{(0)}$. Hence

(2) *The ∞^2 Semple $[n-2]$'s in $[y]$ defined by ${}_y[n-2]_{z^{(0)}}$, where $z^{(0)}$ is on $F_2^{N_n-1}$, are the N_{n-2}-secant spaces of the ∞^{n+1} curves $S^{N_n-1}(\zeta)$ on $Q^n_{N_n}$. Each curve $S^{N_n-1}(\zeta)$ has ∞^1 Semple N_{n-2}-secant spaces corresponding to points $z^{(0)}$ on $Q^{N_n-1}(\zeta)$, or to points $x^{(0)}$ on $Q^{n+1}(\zeta)$. Each Semple $[n-2]$ is N_{n-2}-secant to the ∞^n curves $S^{N_n-1}(\zeta)$ for which the corresponding curves $Q^{N_n-1}(\zeta)$ are on $z^{(0)}$, or the corresponding curves $Q^{n+1}(\zeta)$ are on $x^{(0)}$.*

Some of the correspondences dealt with above have already been obtained analytically. Thus the equation (2) of §4 yields, for $x = x^{(0)}$, the equation $(z^{(0)}\zeta)$ of the corresponding point on the White surface; and for the section Q^{N_n-1} of this surface by the prime ζ, the corresponding curve $Q^{n+1}(\zeta)$ on $P^2_{N_n}$. When bordered as in (2) of §6, it yields the corresponding point $y^{(0)}$ on the corresponding curve $S^{N_n-1}(\zeta)$. For the particular point p_h of $P^2_{N_n}$, it yields (cf. §6, (4)) the line π_h of the double-N_n configuration π_h, κ_h on the White surface, and the corresponding point q_h of $Q^n_{N_n}$.

The matrices in y and z can be bordered similarly to obtain analogous results. Thus

$$(3)\qquad \begin{vmatrix} \alpha_i(\beta y)\gamma_k \\ \zeta_k, \zeta_k', \cdots, \zeta_k^{(n-2)} \end{vmatrix} = 0$$

is, for given y, the equation satisfied by the $(\zeta, \zeta', \cdots, \zeta^{(n-2)})$'s = [2]'s which cut the ${}_z[n-2]_y$ which is N_{n-2}-secant to $F_2^{N_{n-1}}$ or to any $Q^{N_{n-1}}$ section of it by a prime on ${}_z[n-2]_y$. The coefficients of the various coördinates of such a [2] are the linearly independent cubic primals on $Q_{N_n}^n$. When y is at q_h of $Q_{N_n}^n$, it vanishes identically, and, bordered again, it factors as follows:

$$(4)\qquad \begin{vmatrix} \alpha_i(\beta y)\gamma_k & \xi_i \\ \zeta_k, \zeta_k', \cdots, \zeta_k^{(n-1)} & 0 \end{vmatrix} = (p_h\xi)\cdot(\kappa_h\zeta\zeta'\cdots\zeta^{(n-1)}),$$

where κ_h is the $[n-1]$ opposite π_h in the double-N_n configuration and p_h is in $P_{N_n}^2$.

Similarly,

$$(5)\qquad \begin{vmatrix} \alpha_i(\gamma z)\beta_j \\ \eta_j, \eta_j', \cdots, \eta_j^{(n-3)} \end{vmatrix} = 0$$

is, for given z, the equation of the $(\eta, \eta', \cdots, \eta^{(n-3)})$'s = [2]'s which cut ${}_y[n-3]_z$. The coefficients of this equation are the cubic primals on the White surface. When z is a point $z^{(0)}$ of this surface, (5) vanishes identically, and, bordered again, it factors as follows:

$$(6)\qquad \begin{vmatrix} \alpha_i(\gamma z^{(0)})\beta_j & \xi_i \\ \eta_j, \eta_j', \cdots, \eta_j^{(n-2)} & 0 \end{vmatrix} = (x^{(0)}\xi)\cdot(\rho^{(0)}\eta\eta'\cdots\eta^{(n-2)}),$$

where $x^{(0)}$ is the point on [2] which corresponds to $z^{(0)}$, and where $\rho^{(0)}$ is the Semple $[n-2]$ corresponding to $x^{(0)}$ or $z^{(0)}$ as described above.

8. **Properties of the Semple congruence.** One of the most striking properties of the Semple congruence of ∞^2 $[n-2]$'s is expressed by the following theorem which not only gives a solution of the classic problem of projectivity, but also states a fundamental property of the set $Q_{N_n}^n$.

(1) *If $\rho^{(0)}$ is the Semple $[n-2]$ in the space $[n]$ of y which corresponds to $x^{(0)}$ in the plane, then the N_n primes of the pencil on $\rho^{(0)}$ to the points of $Q_{N_n}^n$ are projective to the N_n lines of the pencil on $x^{(0)}$ to the points of $P_{N_n}^2$.*

To prove this let, as in (13) of §5, the pencil $\lambda Q^{n+1}(\zeta) + \lambda' Q^{n+1}(\zeta')$ have base points $R_{N_{n-1}}^2$ outside $P_{N_n}^2$. Then, according to this theorem, n-ic curves on $R_{N_{n-1}}^2$ map the plane into a White surface $F_2^{N_{n-2}}$, and the pencil into a pencil of curves $S^{N_{n-1}}(\lambda\zeta + \lambda'\zeta')$ on $F_2^{N_{n-2}}$ and through $Q_{N_n}^n$, the map of $P_{N_n}^2$. If $x^{(0)}$ is one point of $R_{N_{n-1}}^2$, there is a unique $(n-1)$-ic G on the remaining points

of $R^2_{N_{n-1}}$. Thus, if L is a line of the pencil on $x^{(0)}$, we have in LG a pencil of the mapping system, and therefore in $[y]$ a corresponding pencil of primes on an $[n-2]$, ρ. The curve G, meeting the curves $Q^{n+1}(\zeta)$, $Q^{n+1}(\zeta')$ in N_{n-2} points outside $R^2_{N_{n-1}}$, maps into a curve in ρ which meets the curves $S^{N_{n-1}}(\lambda\zeta + \lambda'\zeta')$ in N_{n-2} points, whence ρ is a Semple $[n-2]$ which is also a space κ of the double-N_{n-1} configuration on $F_2^{N_{n-2}}$. Hence the pencil of primes on ρ to the points of $Q^n_{N_n}$ is projective to the pencil of lines L on $x^{(0)}$ to the points of $P^2_{N_n}$. There remains only to prove that ρ is the Semple $[n-2]$ which corresponds to $x^{(0)}$, i.e., that the points y on ρ satisfy

$$\text{(a)}\ \alpha_i(\beta y)(\gamma z^{(0)}) = 0, \qquad \text{(b)}\ (\alpha x^{(0)})(\gamma z^{(0)})\beta_j = 0.$$

The point $x^{(0)}$ of $R^2_{N_{n-1}}$ corresponds to a $z^{(0)}$ on $F_2^{N_{n-1}}$ for which $(\zeta z^{(0)}) = (\zeta' z^{(0)}) = 0$. These two linear conditions on z together with $T(2, n, n+1) = (\alpha x)(\beta y)(\gamma z) = 0$ yield a $T'(2, n, n-1) = (\alpha' x)(\beta' y)(\gamma' z') = 0$, where z' is in the $[n-1] = (\zeta\zeta')$. This has N_{n-1} pairs $x^{(0)}$, $z^{(0)} = R^2_{N_{n-1}}$, $R'^{n-1}_{N_{n-1}}$ neutral for y as in (b). Thus T' defines the White surface $F_2^{N_{n-2}}$ in $[y]$, and $z^{(0)}$ defines as in (a) a space κ of the configuration on $F_2^{N_{n-2}}$. Hence the Semple space ρ is that which corresponds to $x^{(0)}$.

An incidental result obtained above, transferred from $F_2^{N_{n-2}}$ to the White surface $F_2^{N_{n-1}}$, is the following:

(2) *If p_h, q_h is a neutral pair of $T(2, n, n+1)$, and if κ_h is the corresponding $[n-1]$ opposite π_h on $F_2^{N_{n-1}}$, if also $x^{(s)}$, $y^{(s)}$ $(s = 1, 2, \cdots)$ is any set of corresponding pairs of points on $[2]$ and $F_2^{N_{n-1}}$ respectively, then the pencil of lines on p_h to the points $x^{(s)}$ is projective to the pencil of primes ζ on κ_h to the points $z^{(s)}$.*

Another result, perhaps worth mentioning, is

(3) *If a pencil of curves $Q^{n+1}(\zeta)$ on $P^2_{N_n}$ has further base points $R^2_{N_{n-1}}$, then the N_{n-1} Semple $[n-2]$'s corresponding to the points $R^2_{N_{n-1}}$ are the N_{n-1} spaces κ of a double-N_{n-1} configuration of a White surface $F_2^{N_{n-2}}$ on $Q^n_{N_n}$. The $[n-1]$ in $[z]$ determined by this pencil cuts $F_2^{N_{n-1}}$ in $[z]$ in a set of points $Z'^{n-1}_{N_{n-1}}$ such that $R^2_{N_{n-1}}$, $Z'^{n-1}_{N_{n-1}}$ are the neutral pairs x, z' of a $T'(2, n, n-1) = (\alpha' x)(\beta' y)(\gamma' z')$, z' being in the $[n-1] = (\zeta, \zeta')$.*

Thus the generic section of $F_2^{N_{n-1}}$ by an $[n-1]$ is not a generic set of points if $n > 3$. These points are in fact a set $Q^{n-1}_{N_{n-1}}$ for a generic planar $P^2_{N_{n-1}}$.

We have seen, in connection with (1) of §7, that the ∞^n sets $C^2_{N_{n-2}}$ which with $P^2_{N_n}$ carry a net of curves $Q^{n+1}(\zeta)$ are in birational correspondence with the ∞^n points y of $[y]$, and, in connection with (6) of §7, that the ∞^2 points $x^{(0)}$ of $[x]$ are in birational correspondence with the ∞^2 Semple $[n-2]$'s, in $[y]$. We prove now

(4) *If $x^{(0)}$ corresponds to $\rho^{(0)}$, and if $C^2_{N_{n-2}}$ corresponds to y, then y is a point of $\rho^{(0)}$ if $x^{(0)}$ is a point of $C^2_{N_{n-2}}$, and conversely.*

For, if y is given, the N_{n-2} points $z^{(0)}$ which correspond on $F_2^{N_{n-1}}$ to the points $x^{(0)}$ of $C^2_{N_{n-2}}$, and the corresponding points $x^{(0)}$, are given by the equations $\alpha_i(\beta y)(\gamma z^{(0)}) = 0$ and $(\alpha x^{(0)})(\gamma z^{(0)})\beta_j = 0$. On the other hand, $\rho^{(0)}$ as deter-

mined by $x^{(0)}$ is also given by $(\alpha x^{(0)})(\gamma z^{(0)})\beta_j = 0$, $\alpha_i(\beta y)(\gamma z^{(0)}) = 0$. As an immediate consequence we have

(5) *The order of the Semple congruence is* N_{n-2}.

The N_{n-2} Semple $[n - 2]$'s on a point $y^{(0)}$ are not a generic set if $n > 3$. Indeed, if the space $[z'] = {}_z[n - 2]_{y^{(0)}}$ cuts $F_2^{N_{n-1}}$ in the set of points $Z'^{n-2}_{N_{n-2}}$ which corresponds to the set $C^2_{N_{n-2}}(y^{(0)})$ in $[x]$, and if the space $[n - 1] = [y']$ is the projection of the $[n] = [y]$ from $y^{(0)}$, we have the theorem:

(6) *The* ∞^2 *curves* $S^{N_{n-1}}(\zeta)$ *on* $y^{(0)}$ *are projected from* $y^{(0)}$ *into a net of curves of order* $N_{n-1} - 1$ *on a White surface* $F_2^{N_{n-3}}$ *in* $[y']$ *which contains the set* $Q'^{n-1}_{N_n}$, *the projection of* $Q^n_{N_n}$ *from* $y^{(0)}$. *The* N_{n-2} *Semple* $[n - 2]$*'s on* $y^{(0)}$ *project into the* $[n - 3]$*'s,* κ, *of* $F_2^{N_{n-3}}$. *The White surface* $F_2^{N_{n-3}}$ *is defined by a* $T'(2, n - 1, n - 2) = T'(x, y', z')$ *with sets* $P^2_{N_{n-2}}$, $Q^{n-2}_{N_{n-2}} = C^2_{N_{n-2}}(y^{(0)})$, $Z'^{n-2}_{N_{n-2}}$ *of points* x, z' *neutral for* y'.

Thus while the set of points $P^2_{N_{n-2}} = C^2_{N_{n-2}}(y^{(0)})$ is still a generic set, the set of points $Z'^{n-2}_{N_{n-2}}$ and the set of Semple $[n - 2]$'s on $y^{(0)}$ are special if $n > 3$.

The theorem may be proved by the mapping described in (13) of §5 when the set $R^2_{N_{n-1}}$ is specialized into a set $C^2_{N_{n-2}}(y^{(0)})$ and a set L_n of n points on a line L. The line L maps into the point $y^{(0)}$, and the members of the mapping system containing L constitute the $(n - 1)$-ics on $C^2_{N_{n-2}}(y^{(0)})$. However, we may also think of z in $T = 0$ as restricted to the ${}_z[n - 2]_{y^{(0)}}$ defined by the equations $\alpha_i(\beta y^{(0)})(\gamma z) = 0$. For z so restricted $T = 0$ has a neutral point $y^{(0)}$. If y in $[n]$ is $\lambda y^{(0)} + y'$, where y' is in an $[n - 1]$ not on $y^{(0)}$, then in the spaces $[y']$, $[z']$, $T = 0$ reduces to a $T'(2, n - 1, n - 2) = T'(x, y', z')$, which defines in $[y']$ a White surface $F_2^{N_{n-3}}$. The pairs p_h, q_h of points x, y neutral for z in $T = 0$ are for $T' = 0$ ordinary corresponding pairs x, y' on $[2]$, $F_2^{N_{n-3}}$, whereas the pairs $x^{(0)}$, $z^{(0)}$ in $C^2_{N_{n-2}}(y^{(0)})$, $Z'^{n-2}_{N_{n-2}}$ are the pairs neutral for y' in $T' = 0$. Such a point $z^{(0)}$ corresponds in $T' = 0$ to a space $\kappa = [n - 3]$ of $F_2^{N_{n-3}}$, which, dilated from $y^{(0)}$, is the Semple $[n - 2]$ of $T = 0$ corresponding to $z^{(0)}$.

If q_h is a point of $Q^n_{N_n}$, the equations $\alpha_i(\beta q_h)(\gamma z) = 0$ are dependent with multipliers p_h, and they define a space ${}_z[n - 1]_{q_h} = \kappa_h$ of the double-N_n configuration on $F_2^{N_{n-1}}$. This κ_h cuts $F_2^{N_{n-1}}$ in a curve $Q_h^{N_{n-2}}$ which corresponds in $[x]$ to the curve Q^n_h of order n on all of the points of $P^2_{N_n}$ except p_h. If z is restricted to the space $[z'] = {}_z[n - 1]_{q_h}$, then q_h is a neutral point in $[y]$ for $T = 0$. If then the space $[y]$ is projected from q_h into a space $[y'] = [n - 1]$, T becomes a trilinear form $T'(2, n - 1, n - 1) = T'(x, y', z')$. The three equations above show that the Semple $[n - 2]$'s corresponding to points z' on $Q_h^{N_{n-2}}$ are on q_h, and thus they project from q_h into $[n - 3]$'s in $[y']$. In $T' = 0$ we have a situation like that of (2) in §6. Hence the birationally related curves $Q_h^{N_{n-2}}$, Q^n_h are birationally related to a curve $S'^{N_{n-2}}_h$ in $[y']$ on the projection $Q'^{n-1}_{N_n-1}$ from q_h so that the points z' on $Q_h^{N_{n-2}}$ correspond to the N_{n-3}-secant spaces of $S'^{N_{n-2}}_h$ which arise from the Semple $[n - 2]$'s on q_h. Hence

(7) *The* ∞^1 *Semple* $[n - 2]$*'s on a point* q_h *of* $Q^n_{N_n}$ *project from* q_h *into the* ∞^1 N_{n-3}*-secant* $[n - 3]$*'s of a curve* $S'^{N_{n-2}}_h$ *on the projection* $Q'^{n-1}_{N_n-1}$ *of the remaining*

points of $Q^n_{N_n}$. *This curve* $S'^{N_n-2}_h$ *is birationally related to the curves* $Q^{N_n-2}_h$ *on* $F^{N_n-1}_2$ *in* κ_h, *and* Q^n_h *in* $[x]$. $Q^{N_n-2}_h$ *is the map of* Q^n_h *by curves* $Q^{n+1}(\zeta)$, *and* $S'^{N_n-2}_h$ *is the map of* Q^n_h *by* $(n-1)$*-ics on a set of the first mapping, the points other than* p_h *on* Q^n_h *passing into* $Q'^{n-1}_{N_n-1}$.

We now examine the Semple $[n-2]$'s corresponding to points z on a line π_h of $F^{N_n-1}_2$ and, with respect to them, prove that:

(8) *If* B_h *is the quadric on all of the points of* $Q^n_{N_n}$ *except* q_h, *then* B_h *has a nodal space* $[n-4]$, *say* $N[n-4]_h$. *It is therefore twice ruled with* ∞^1 *spaces* $[n-2]$. *One of these two rulings is the locus of Semple* $[n-2]$*'s which correspond to points* z *on* π_h *of* $F^{N_n-1}_2$, *or to directions about* p_h *in* $[x]$.

For, if $x^{(0)}$ in x corresponds to $z^{(0)}$ on $F^{N_n-1}_2$, the point $z^{(0)}$ defines a Semple $[n-2]$ because the equations $\alpha_i(\beta y)(\gamma z^{(0)}) = 0$ are dependent with multipliers $x^{(0)}$. Now, if $z^{(0)}$ is on π_h, then $x^{(0)}$ is p_h, and the coefficients of dependence are constant as $z^{(0)}$ travels along the line $\pi_h = z^{(0)\prime} + tz^{(0)\prime\prime}$. Hence one of the three equations may be discarded, and the other two yield two projective pencils of primes in $[n]$ which generate a quadric of the type described in (8). Since π_h cuts each curve $Q^{N_n-2}_{h'}$ (cf. (7)) in a point if $h' \neq h$, the corresponding $[n-2]$ is on $q_{h'}$, and the quadric is B_h.

Thus we have, for $n \geqq 4$, a new configuration in $[n]$ of N_n nodal spaces $N[n-4]_h$. For $n = 4$, this is a new set of 15 points defined by Q^4_{15}. For $n = 3$, the nodal spaces are non-existent, but it is still true that one set of generators of the ten quadrics B_h is a set of Semple lines. For $n = 2$, the Semple $[n-2]$'s are the points y, and the quadrics B_h are conics on 5 of the points Q^2_6.

If $n = 3$, the Semple congruence of lines has no other singular points than the triple points Q^3_{10}. However, for $n \geqq 4$ it has also these loci of double singular points $N[n-4]_h$.

Reverting to (1) we observe that

(9) *Of the two ordered point sets* $P^2_{N_n}$, $Q^n_{N_n}$, $2n+1$ *pairs* p_h, q_h *can be chosen at random, the remaining* N_{n-2} *pairs being then uniquely determined.*

First, it is clear that two such sets of $2n+1$ pairs have respectively $2(2n-3)$, $n(n-1)$ absolute constants, and thus the two have as many as $P^2_{N_n}$ itself. Secondly, the algebraic identity (4) of §5 has already been used to determine, from the given $2n+1$ pairs $(q_h\eta)\cdot(p_h\xi)$, the $T(2, n, n+1)$ for which they are neutral pairs. The remaining neutral pairs are then fixed. Thirdly, the $2n+1$ given pairs are sufficient to determine the Semple $\rho^{(0)}$ in $[n]$ defined by $x^{(0)}$ in [2]. For, the projectivity given in (1) imposes $2(n-2)$ conditions on $\rho^{(0)}$, which are sufficient to determine it uniquely. Indeed, the Semple congruence may be obtained from the $2n+1$ given pairs by using the method applied in (6). If the $2n+1$ given points q_h are projected from $y^{(0)}$ into $2n+1$ points q'_h in an $[n-1] = [y']$, then the $2n+1$ pairs p_h, q'_h are neutral for z' in a $T'(2, n-1, n-2) = T'(x, y', z')$. This defines a White surface F^{N_n-3} in $[y']$ on the points q'_h, these corresponding to p_h in $[x]$. The

N_{n-2} spaces $\kappa = [n-3]$ of this surface, dilated from $y^{(0)}$, yield the N_{n-2} Semple $[n-2]$'s on $y^{(0)}$. With the Semple congruence thus defined, its singular points $Q^n_{N_n}$ are all determined.

9. **Degenerate curves of the birationally related systems,** $Q^{n+1}(\zeta)$, $Q^{N_{n-1}}(\zeta)$, $S^{N_{n-1}}(\zeta)$. The curve $Q^{n+1}(\zeta)$ is mapped on $Q^{N_{n-1}}(\zeta)$ by curves $Q^{n+1}(\zeta')$, and on $S^{N_{n-1}}(\zeta)$ by n-ic curves on a further set $R^2_{N_{n-1}}$ cut out on $Q^{n+1}(\zeta)$ by some curve $Q^{n+1}(\zeta')$. If then $Q^{n+1}(\zeta)$ itself is composite, the curves $Q^{N_{n-1}}(\zeta)$ and $S^{N_{n-1}}(\zeta)$ will usually, though not always, be composite. However, the composite curves $Q^{n+1}(\zeta)$ will, if $n > 3$, be confined to two types of pencils each with a fixed part. These are first the fixed curve Q^n_h on all the points of $P^2_{N_n}$ except p_h together with the pencil of lines on p_h, and secondly the fixed line p_h, $p_{h'}$ together with the pencil of curves $Q^n_{h,h'}$ on all of the points of $P^2_{N_n}$ except p_h, $p_{h'}$. If $n = 3$, there will also be a finite number of pairs of conics on P^2_{10}, and, if $n = 2$, a finite number of triangles on P^2_6.

However, the curves $Q^{N_{n-1}}(\zeta)$ and $S^{N_{n-1}}(\zeta)$ may degenerate without $Q^{n+1}(\zeta)$ becoming composite. Consider a curve $Q^{n+1}(\zeta) = (p_1^2 p_2 \cdots p_{N_n})^{n+1} = Q$. Its genus is reduced by one because of the node at p_1. Since it no longer goes through p_1 with a definite direction (as does the generic $Q^{n+1}(\zeta)$), we think of it as made up of the directions at p_1, say p_1^*, and of the actual curve Q, these two having in common two directions or points. Under the first mapping, p_1^* becomes a line π_1 on $F_2^{N_{n-1}}$, and Q becomes a curve of order $N_{n-1} - 1$ which meets π_1 in two points. Thus we secure again the intersection of $F_2^{N_{n-1}}$ by a prime ζ on π_1. The set $R^2_{N_{n-1}}$ must now contain the point p_1 so that $S^{N_{n-1}}(\zeta)$ also is made up of a curve of order $N_{n-1} - 1$ and a bisecant. The types of degeneration of this sort tend to increase with increasing n since the multiplicities at one or more points of $P^2_{N_n}$ may increase.

The N_{n-2}-secant $[n-2]$'s of each of the curves $Q^{N_{n-1}}(\zeta)$, $S^{N_{n-1}}(\zeta)$ are related to the points of the other. Thus, when decompositions occur, these families of ∞^1 $[n-2]$'s also degenerate. Since all of the $[n-2]$'s for curves $S^{N_{n-1}}(\zeta)$ are Semple $[n-2]$'s, a great deal of specific information about this congruence can be obtained from a study of the decompositions mentioned for specific values of n. The case $n = 3$ is of unusual interest in this connection since the $Q^6(\zeta)$, $S^6(\zeta)$ define a cubic Cremona transformation between the two [3]'s of ζ and y.

10. **Conclusion.** We hope in a following paper to show that if $P^2_{N_n}$ is the set of nodes of a ternary rational curve of order $n+3$, then there exist two trilinear forms $T(2, n, n+1)$ and $T'(2, n, n+1)$, the one with neutral sets $P^2_{N_n}$, $Q^n_{N_n}$ and the other with neutral sets $P'^2_{N_n}$, $Q^n_{N_n}$, the sets $P^2_{N_n}$, $P'^2_{N_n}$ being projectively distinct, but the set $Q^n_{N_n}$ being the same for both. In other words, the geometry of these two forms is superposed in the space $[n]$ of y. For this reason we have been interested above mainly in the set $Q^n_{N_n}$ and the Semple congruence.

UNIVERSITY OF ILLINOIS.

CONDITIONS ON THE NODES OF A RATIONAL PLANE CURVE

By Arthur B. Coble

1. Introduction. The rational curve in the plane of order $n + 3$ with rational parameter t, say $\rho_2^{n+3}(t)$, has a set of $N_n = \binom{n+2}{2}$ nodes at points p_h, say the set $P^2_{N_n}$. If $n < 3$, these nodes may be taken generically and the rational curve exists. For $n = 3$, however, the ten nodes are subject to three conditions first noticed by Valentiner in 1881.[1] For $n > 3$ similar conditions have not been obtained. It is the purpose of this paper to obtain necessary conditions on the set $P^2_{N_n}$ for generic n. For this we define in §4 a set of N_n "nodular" points $P'^2_{N_n}$, which is shown to be not projective to $P^2_{N_n}$ in general. In §5 a set of N_n "catalectic" points $Q^n_{N_n}$ appears in the space $[n]$ of the conjugate rational envelope $r_n^{n+3}(t)$. We prove in §7 that there is a trilinear form $T(2, n, n + 1) = (\alpha x)(\beta y)(\gamma z)$ with pairs $x, y = p_h, q_h$ drawn from $P^2_{N_n}$, $Q^n_{N_n}$ which are neutral for z in $T = 0$; and in §6 that there is a trilinear form $T'(2, n, n + 1)$ with neutral pairs $x', y = p'_h, q_h$ drawn from $P'^2_{N_n}$, $Q^n_{N_n}$ which are neutral for z in $T' = 0$. The identity of the set $Q^n_{N_n}$ for the non-projective sets $P^2_{N_n}$, $P'^2_{N_n}$ yields the conditions desired. These conditions are sufficient for $n = 3$. We obtain the forms T, T' from the rational curve and its conjugate rational envelope by using certain special coördinate systems developed in §§2, 3. The pertinent theory of such forms is given in a recent paper.[2]

2. A special coördinate system in $[n + 1]$. Two binary forms of order $n + 1$,

$$(1)\qquad \begin{aligned} (\alpha t)^{n+1} &= (\alpha_0 t_0 + \alpha_1 t_1)^{n+1} = \sum_j \binom{n+1}{j} a_j t_0^{n+1-j} t_1^j, \\ (\beta t)^{n+1} &= (\beta_0 t_0 + \beta_1 t_1)^{n+1} = \sum_j \binom{n+1}{j} b_j t_0^{n+1-j} t_1^j \quad (j = 0, \cdots, n + 1), \end{aligned}$$

have a bilinear invariant

$$(2)\qquad (\alpha\beta)^{n+1} = \sum_j (-1)^j \binom{n+1}{j} a_j b_{n+1-j}.$$

If then we put these forms into correspondence with respectively the primes ζ and the points z of a space $[n + 1]$ by setting

$$(3)\qquad \zeta_j = a_j, \qquad z_j = (-1)^j \binom{n+1}{j} b_{n+1-j},$$

Received July 16, 1940.

[1] For references cf. A. B. Coble, *The ten nodes of the rational sextic and of the Cayley symmetroid*, Amer. Jour. of Math., vol. 41(1919), pp. 243–265; pp. 251–254.

[2] A. B. Coble, *Trilinear forms*, this Journal, vol. 7(1940), pp. 380–395.

the incidence condition of prime ζ and point z is expressed by the vanishing of this invariant,

$$(\zeta z) = (\alpha\beta)^{n+1} = 0. \tag{4}$$

If the two binary forms are perfect powers, say

$$\begin{aligned}(\alpha t)^{n+1} &= (tr)^{n+1} = (t_0 r_1 - t_1 r_0)^{n+1},\\ (\beta t)^{n+1} &= (ts)^{n+1} = (t_0 s_1 - t_1 s_0)^{n+1},\end{aligned} \tag{5}$$

so that the corresponding primes ζ and points z are those of a rational norm-curve in $[n + 1]$, namely,

$$\zeta_j = (-1)^j r_0^j r_1^{n+1-j}, \qquad z_j = (-1)^{n+1} \binom{n+1}{j} s_0^{n+1-j} s_1^j, \tag{6}$$

the incidence condition

$$(\zeta z) = (rs)^{n+1} = (r_0 s_1 - r_1 s_0)^{n+1} = 0 \tag{7}$$

shows that the primes ζ and points z belong to the *same* rational norm-curve N^{n+1} in $[n + 1]$.

Since $(\alpha s)^{n+1}$ is the bilinear invariant of $(\alpha t)^{n+1}$, $(st)^{n+1}$, the roots of $(\alpha t)^{n+1}$ determine the parameters s of points of N^{n+1} on ζ; since $(\beta r)^{n+1}$ is the bilinear invariant of $(tr)^{n+1}$, $(\beta t)^{n+1}$, the roots of $(\beta t)^{n+1}$ determine the parameters r of primes on z. The coefficients a and b may be expressed in terms of symmetric functions of these $n + 1$ roots, these expressions being merely the polarized forms of (6).

A primal of order k in z, say $(\lambda z)^k = 0$, becomes, when z is replaced from (3) and $(\beta t)^{n+1}$ is factored into $(t_1 t) \cdot (t_2 t) \cdot \cdots \cdot (t_{n+1} t)$, a symmetric form in the binary variables $t_1, t_2, \cdots, t_{n+1}$, of degree k in each, say

$$\sum_{n+1}^{k} = (\lambda_1 t_1)^k (\lambda_2 t_2)^k \cdots (\lambda_{n+1} t_{n+1})^k = 0. \tag{8}$$

Conversely, such a symmetric form becomes, when the elementary symmetric combinations are replaced by coördinates z, a primal of order k in $[n + 1]$. This primal of order k in $[n + 1]$ we call the *parametric primal* attached to the symmetric form $\sum_{n+1}^{k}$.

However, $\sum_{n+1}^{k}$ may be interpreted in a quite different fashion. If, in a space $[k]$, there is a norm-curve N^k with points x defined analogously to points z on N^{n+1} in (6), then the k-ic combinations of t_1 can be replaced by a point $x^{(1)}$ on N^k, those of t_2 by a point $x^{(2)}$ on N^k, etc. Then $\sum_{n+1}^{k}$ becomes a form which is linear and symmetric in $x^{(1)}, \cdots, x^{(n+1)}$, and which therefore is the completely polarized form of a primal of order $n + 1$ in $[k]$. This primal of order $n + 1$ in $[k]$ we call the *apolarity primal* attached to the symmetric form $\sum_{n+1}^{k}$.

This coördinate system, based on the norm-curve N^{n+1} with $n(n+4)$ constants, and on the distribution of the binary parameter t on it with 3 constants, is obviously generic. The duality by which a binary form represents either a prime or a point is that set up by taking α, β in (2) as points z, z' or as primes ζ, ζ'. This correlation is a polarity or a null-system according as n is odd or even.

3. **A special coördinate system in** $[n] = [2k+1]$, $[2k]$. Two double binary forms in digredient variables τ, t, with respective orders 1, k in these variables,

$$(1)\qquad \begin{aligned} (m\tau)(\mu t)^k &= (m_0\tau_0 + m_1\tau_1)(\mu_0 t_0 + \mu_1 t_1)^k = \sum_{i,j} \binom{k}{j} m_{i,j}\tau_i t_0^{k-j} t_1^j, \\ (m'\tau)(\mu' t)^k &= (m'_0\tau_0 + m'_1\tau_1)(\mu'_0 t_0 + \mu'_1 t_1)^k = \sum_{i,j} \binom{k}{j} m'_{i,j}\tau_i t_0^{k-j} t_1^j \\ &\qquad (i = 0, 1;\ j = 0, \cdots, k) \end{aligned}$$

have a bilinear invariant under digredient transformation,

$$(2)\qquad (mm')(\mu\mu')^k = \sum_{i,j} (-1)^{i+j} \binom{k}{j} m_{i,j} m'_{1-i,\,k-j}\,.$$

If then we put these forms into correspondence with respectively the primes η and the points y of a space $[2k+1]$ by setting

$$(3)\qquad \eta_{i,j} = m_{i,j}, \qquad y_{i,j} = (-1)^{i+j}\binom{k}{j} m'_{1-i,k-j},$$

the incidence condition of prime η and point y is expressed by the vanishing of this invariant,

$$(4)\qquad (\eta y) = (mm')(\mu\mu')^k = 0.$$

If each of the two double forms is a product of perfect powers, say

$$(5)\qquad \begin{aligned} (m\tau)(\mu t)^k &= (\tau\rho)\cdot(tr)^k = (\tau_0\rho_1 - \tau_1\rho_0)\cdot(t_0 r_1 - t_1 r_0)^k, \\ (m'\tau)(\mu' t)^k &= (\tau\rho')\cdot(tr')^k = (\tau_0\rho'_1 - \tau_1\rho'_0)\cdot(t_0 r'_1 - t_1 r'_0)^k, \end{aligned}$$

then the corresponding primes η and points y are those of rational surfaces S, S' in $[2k+1]$, namely,

$$(6)\qquad \eta_{i,j} = (-1)^{i+j}\rho_{1-i} r_0^j r_1^{k-j}, \qquad y_{i,j} = (-1)^{k+1}\binom{k}{j}\rho'_i r_0'^{k-j} r_1'^{j},$$

the incidence condition of the prime η and point y in (6) being

$$(7)\qquad (\eta y) = (\rho\rho')\cdot(rr')^k = (\rho_0\rho'_1 - \rho_1\rho'_0)\cdot(r_0 r'_1 - r_1 r'_0)^k.$$

The locus S' of the point $y_{i,j}$, which in (6) depends on the two binary parameters $\rho'_0:\rho'_1$ and $r'_0:r'_1$, is the surface known as a general rational scroll in $[n] = [2k+1]$.[3] For fixed r' and variable ρ', the point $y_{i,j}$ runs over a generator

[3] T. G. Room, *The Geometry of Determinantal Loci*, Cambridge University Press, 1938. Cf. p. 237, 11.8.4.

$g(r')$; for fixed ρ' and variable r', the point $y_{i,j}$ runs over a directrix $N^k(\rho')$. Through each point of S' there is one generator and one directrix.

The locus S of primes, defined in (6), is the dual in $[n] = [2k + 1]$ of S'. A prime of S, according to (7), cuts S' in a directrix $N^k(\rho)$ and a generator $g(r)$ repeated k times. We may define S, S' to be dual forms of the *same* locus.

It is clear from (1), (3), and (6) that we may interpret $(m\tau')(\mu t')^k = 0$ as the condition that the prime $m_{i,j}$ in $[n]$ is on the point τ', t' of S'; and similarly, $(m'\tau)(\mu' t)^k = 0$ as the condition that the point $m'_{i,j}$ in $[n]$ is on the prime τ, t of S.

A particular case of this representation will be needed later when n is even. If attention is confined to those forms $(m'\tau)(\mu' t)^k$ which satisfy the condition

$$(8) \qquad (m'\tau^{(0)})(\mu' t^{(0)})^k = 0,$$

where $\tau^{(0)}$, $t^{(0)}$ is a fixed pair of digredient binary parameters, then obviously we are dealing with points $m'_{i,j}$ which lie on the prime $= [2k]$ of S defined by the parameters $\tau^{(0)}$, $t^{(0)}$.

4. **The perspective curves and nodular points of $\rho_2^{n+3}(t)$.** Let $\rho_2^{n+3}(t)$ be given by the equation $(a\xi)(\alpha t)^{n+3} = 0$, t being a binary parameter and ξ a line coördinate in the space $[x] = [2]$. A ternary rational envelope of class ν, $(\beta x)(\gamma t)^\nu = 0$, is said to be *perspective* to $\rho_2^{n+3}(t)$ if tangent t of the envelope is on point t of $\rho_2^{n+3}(t)$, i.e., if

$$(a\beta)(\alpha t)^{n+3}(\gamma t')^\nu \equiv (tt')\cdot(\lambda t)^{n+2}(\mu t')^{\nu-1},$$

or

$$(1) \qquad (a\beta)(\alpha t)^{n+3}(\gamma t)^\nu \equiv 0.$$

This imposes $\nu + n + 4$ linear homogeneous conditions on the $3(\nu + 1)$ coefficients of the envelope, whence[4] $\rho_2^{n+3}(t)$ has $\infty^{2\nu-n-2}$ perspective envelopes of class ν. Thus, if n is odd, $\rho_2^{n+3}(t)$ has ∞^1 perspective envelopes of class $\nu = \frac{1}{2}(n + 3)$. Any two of these generate $\rho_2^{n+3}(t)$ in the sense that tangents t of the two envelopes meet in point t of $\rho_2^{n+3}(t)$. However, if n is even, $\rho_2^{n+3}(t)$ has a unique perspective envelope of class $\nu = \frac{1}{2}(n + 2)$, and ∞^2 of class $\frac{1}{2}(\nu + 4)$. Then $\rho_2^{n+3}(t)$ is generated by the one of class $\frac{1}{2}(n + 2)$ and any proper one of class $\frac{1}{2}(n + 4)$.

If $2\nu - n - 2$ is negative, then $n + 2 - 2\nu$ is the number of conditions imposed on $\rho_2^{n+3}(t)$ itself in order that it may have a perspective curve of class as low as ν. In particular we see that

(2) *The condition that $\rho_2^{n+3}(t)$ have a perspective conic imposes $n - 2$ conditions on $\rho_2^{n+3}(t)$.*

[4] A. B. Coble, *Symmetric binary forms and involutions*, III, Amer. Jour. of Math., vol. 32(1910), pp. 355–364. See §15.

Rational curves with a perspective conic are treated in the paper cited in footnote 4 (cf. especially p. 357, (133)). We consider here only such curves as have no perspective envelopes of class less than the limits found above, i.e., $\frac{1}{2}(n+3)$ (n odd), $\frac{1}{2}(n+2)$ (n even), and also such that $n \geqq 3$. Thus we exclude all rational curves with a perspective conic.

In case n is odd, the perspective envelopes of class $\frac{1}{2}(n+3)$ can be given by the generic form

$$(3) \qquad (\pi x)(d\tau)(\delta t)^{\frac{1}{2}(n+3)} = 0 \qquad (n \text{ odd}).$$

Such a form in ternary x, and digredient binary variables τ, t has $3 \cdot 2 \cdot \frac{1}{2}(n+5) - 1 - 8 - 2 \cdot 3 = 3n$ absolute constants. On the other hand, $(a\xi)(\alpha t)^{n+3}$ also has $3(n+4) - 1 - 3 - 8 = 3n$ absolute constants. If two of the perspective envelopes are given by τ and τ', the meet of tangents t of these envelopes, to within the factor $(\tau\tau')$, is

$$(4) \qquad (a\xi)(\alpha t)^{n+3} \equiv (\pi\pi'\xi)(dd')(\delta t)^{\frac{1}{2}(n+3)}(\delta' t)^{\frac{1}{2}(n+3)} = 0.$$

If n is even, we consider the perspective envelopes

$$(3') \qquad (\pi x)(d\tau)(\delta t)^{\frac{1}{2}(n+4)} = 0 \qquad (n \text{ even}),$$

with the subsidiary condition

$$(3'.1) \qquad (\pi x)(d\tau^{(0)})(\delta t^{(0)})^{\frac{1}{2}(n+4)} \equiv 0 \qquad (\text{in } x).$$

As τ varies in (3′) we again have ∞^1 envelopes. However, according to (3′.1), the particular envelope defined by $\tau = \tau^{(0)}$ has the extraneous factor $(tt^{(0)})$ in its parametric equation and thus is actually the unique perspective envelope of class $\frac{1}{2}(n+2)$. Thus the equation of the rational curve generated by the envelopes (3′) is

$$(4') \qquad (a\xi)(\alpha t)^{n+3} \equiv (\pi\pi'\xi)(dd')(\delta t)^{\frac{1}{2}(n+4)}(\delta' t)^{\frac{1}{2}(n+4)}/(tt^{(0)}) = 0.$$

The form (3′) has $3n + 3$ absolute constants, 3 of which are eliminated by (3′.1), leaving $3n$ for $\rho_2^{n+3}(t)$. Again, $\rho_2^{n+3}(t)$ has $3n$ absolute constants and determines its unique perspective $\frac{1}{2}(n+2)$-ic. In (3′) a system (∞^1) of perspective $\frac{1}{2}(n+4)$-ics containing this unique curve is selected, thus adding one absolute constant. But also in (3′.1) the value $\tau = \tau^{(0)}$ is assigned to this curve, and the value $(tt^{(0)})$ is assigned to the extraneous factor, so that the $3n + 3$ constants in (3′) are apparent.

Finally, all of the perspective envelopes of $\rho_2^{n+3}(t)$ are obtained from (3) or (3′) by eliminating τ by using the relations:

$$(5) \qquad (m\tau)(\mu t)^k = 0 \qquad (n \text{ odd}),$$

$$(5') \qquad (m\tau)(\mu t)^k = 0, \qquad (m\tau^{(0)})(\mu t^{(0)})^k = 0 \qquad (n \text{ even}).$$

The resulting systems of perspective envelopes are

$$(6) \qquad (\pi x)(dm)(\delta t)^{\frac{1}{2}(n+3)}(\mu t)^k = 0,$$

$$(6') \qquad (\pi x)(dm)(\delta t)^{\frac{1}{2}(n+4)}(\mu t)^k/(tt^{(0)}) = 0.$$

In both cases we have systems ($\infty^{2\nu-n-2}$) of class ν, the $2\nu - n - 2$ parameters being the coefficients $m_{i,j}$ in (5) and (5').

Let p_h be a node of $\rho_2^{n+3}(t)$ in the nodal set $P_{n+2,2}^2$, and let $(\kappa_h t)^2$ be the quadratic which defines the nodal parameters of p_h on $\rho_2^{n+3}(t)$. Let $K(t)$ be a norm-conic with parameter t in a plane $[x']$. Then $(\kappa_h t)^2$ defines, as in §2, a point $x' = n_h$, this being the meet of the two tangents $(\kappa_h t)^2$ of $K(t)$. We call n_h a *nodule* of $\rho_2^{n+3}(t)$, and the set of nodular points n_h the nodular set $P'^2_{n+2,2}$ of $\rho_2^{n+3}(t)$. Fundamental for our purpose is the theorem:

(7) *If $\rho_2^{n+3}(t)$ does not have a perspective conic, the nodular set $P'^2_{n+2,2}$ in $[x']$ and the nodal set $P^2_{n+2,2}$ in $[x]$ are not projective figures.*

For, if $P'^2_{n+2,2}$ were projective to $P^2_{n+2,2}$, and if it were projected upon $P^2_{n+2,2}$, the norm-conic $K(t)$ used for the construction of $P'^2_{n+2,2}$ would be projected into a conic $K'(t)$ in the plane of $P^2_{n+2,2}$ with the property that tangents $(\kappa_h t)^2$ of $K'(t)$ would meet in the node p_h. The incidence condition of tangent t' of $K'(t)$ and point t of $\rho_2^{n+3}(t)$ would be given by a form $f(t'^2, t^{n+3}) = 0$. Since this would be satisfied for $t' = t$ by each of the $(n + 2)(n + 1)$ nodal parameters, $f(t^2, t^{n+3}) \equiv 0$, and $f(t'^2, t^{n+3})$ would have the factor (tt'). Thus $K'(t)$ would be a perspective conic of $\rho_2^{n+3}(t)$. Conversely, if $\rho_2^{n+3}(t)$ has a perspective conic $K'(t)$, the nodal set $P^2_{n+2,2}$ is also the nodular set referred to $K'(t)$.

5. The conjugate rational curve $r_n^{n+3}(t)$ and its catalectic set $Q^n_{n+2,2}$. With $\rho_2^{n+3}(t)$ given by the equation $(a\xi)(\alpha t)^{n+3} = 0$, we define in a space $[y] = [n]$ the conjugate rational $(n + 3)$-ic locus of primes,

$$(\beta y)(\epsilon t)^{n+3} \equiv E_0(y)t_0^{n+3} + \binom{n+3}{1}E_1(y)t_0^{n+2}t_1 + \binom{n+3}{2}E_2(y)t_0^{n+1}t_1^2 + \cdots = 0, \tag{1}$$

the term "conjugate" denoting that the point sections of the envelope $r_n^{n+3}(t)$ by the points y and the line sections of $\rho_2^{n+3}(t)$ by the lines ξ are given by apolar $(n + 3)$-ics, i.e., that

$$(a\xi)(\alpha\epsilon)^{n+3}(\beta y) \equiv 0 \qquad (\text{in } \xi, y). \tag{2}$$

For given y_h, (1) determines the parameters t of the $n + 3$ primes of $r_n^{n+3}(t)$ on y_h. This $(n + 3)$-ic may have an apolar quadratic $(\kappa_h t)^2$ and be reducible to a sum of powers of the two linear factors of $(\kappa_h t)^2$. Then $(\beta y_h)(\epsilon\kappa_h)^2(\epsilon t)^{n+1} \equiv 0$, and $(\beta y)(\epsilon\kappa_h)^2(\epsilon t)^{n+1}$ contains, for variable y, only n independent $(n + 1)$-ics. Thus a *pencil* $(\kappa t)^{n+1}$ of $(n + 1)$-ics can be found such that $(\beta y)(\epsilon\kappa_h)^2(\epsilon\kappa)^{n+1} \equiv 0$ in y. Hence $(\kappa_h t)^2 \cdot (\kappa t)^{n+1}$ is a pencil of line sections of $\rho_2^{n+3}(t)$, and the fixed factor $(\kappa_h t)^2$ of the pencil is the pair of nodal parameters of a node p_h of $\rho_2^{n+3}(t)$. Conversely, if $(\kappa_h t)^2$ is a pair of nodal parameters, $(\beta y)(\epsilon\kappa_h)^2(\epsilon t)^{n+1}$ has a pencil of apolar $(n + 1)$-ics and therefore *one* member $(\beta y_h)(\epsilon\kappa_h)^2(\epsilon t)^{n+1}$ of the system vanishes identically, and there is a point y_h for which $(\beta y_h)(\epsilon t)^{n+3}$ has the apolar quadratic $(\kappa_h t)^2$. Similarly, $(\beta y_i)(\epsilon t)^{n+3}$ may have an apolar cubic $(\lambda_i t)^3$, and be reducible to a sum of three $(n + 3)$-th powers. Then a *unique* $(\lambda t)^n$ can be found

such that $(\beta y)(\epsilon\lambda_i)^3(\epsilon\lambda)^n \equiv 0$ in y, and therefore $(\lambda_i t)^3\cdot(\lambda t)^n$ is a line section of $\rho_2^{n+3}(t)$. Thus the points y_i lie on a surface whose points are in one-to-one correspondence with the collinear triads of $\rho_2^{n+3}(t)$. If $(\beta y_k)(\epsilon t)^{n+3}$ has an apolar quartic $(\mu_k t)^4$, the system $(\beta y)(\epsilon\mu_k)(\epsilon t)^{n-1}$ contains n linearly independent $(n-1)$-ics. Hence for any quartic $(\mu_k t)^4$ there is a point y_k. Thus, from this point on, the apolarity conditions fall on the point y alone and not also on the apolar form.

The conditions that $(\beta y)(\epsilon t)^{n+3}$ have an apolar quadratic, cubic, etc. are expressed by the requirement that the successive matrices

$$(3) \qquad \Sigma_0 \equiv \begin{vmatrix} E_0 & E_1 & \cdots & E_{n+1} \\ E_1 & E_2 & \cdots & E_{n+2} \\ E_2 & E_3 & \cdots & E_{n+3} \end{vmatrix}, \quad \Sigma_2 \equiv \begin{vmatrix} E_0 & E_1 & \cdots & E_n \\ E_1 & E_2 & \cdots & E_{n+1} \\ E_2 & E_3 & \cdots & E_{n+2} \\ E_3 & E_4 & \cdots & E_{n+3} \end{vmatrix}, \quad \cdots$$

have the ranks 2, 3, $\cdots$. We call the points y for which this requirement is satisfied a "catalectic locus Σ_{2k}", $2k$ being the dimension of the locus, and $\binom{n+2-k}{k+2}$ its order ($k = 0, 1, \cdots, \frac{1}{2}n$). In particular the catalectic locus Σ_0 consists of a set of $\binom{n+2}{2}$ points q_h in $[n]$, say $Q^n_{n+2,2}$, in correspondence with the nodes p_h of $\rho_2^{n+3}(t)$, since $(\beta q_h)(\epsilon t)^{n+3}$ has, as an apolar quadratic, the nodal parameters $(\kappa_h t)^2$ of p_h.

6. **The osculant $(n+1)$-ics of $r_n^{n+3}(t)$ and the first trilinear form** $T'(2, n, n+1)$. If the parametric equation $(\beta y)(\epsilon t)^{n+3}$ of $r_n^{n+3}(t)$ is polarized with respect to t, we obtain families of so-called "osculant envelopes" of $r_n^{n+3}(t)$. In particular, the osculant $(n+1)$-ic envelopes are given by

$$(1) \qquad (\beta y)(\epsilon t_1)(\epsilon t_2)(\epsilon t)^{n+1} = 0,$$

this being the mixed osculant of t_1, t_2 of class $\nu = n+1$. In the form (1), y is a variable in space $[n]$ defined only to within projective transformation, as is the conjugate rational envelope $r_n^{n+3}(t)$ itself in (1) of §5. If, as in §2, we replace t_1, t_2 by the coördinates x' of a point referred to the norm-curve $N^2 = K(t)$, and if further, as in (6) of §2, we replace the $(n+1)$-ic combinations of t by coördinates z' in a space $[n+1]$ referred to a norm-curve N^{n+1}, then the form (1) becomes a trilinear form

$$(2) \qquad T'(2, n, n+1) = (\alpha' x')(\beta' y)(\gamma' z') = 0.$$

Since the norm-curves N^2, N^{n+1} in $[x']$, $[z']$ respectively are arbitrarily chosen, and the parameter systems t are equally arbitrary, the variables x', y, z' in T' are digredient.

We had observed in §5 that $(\beta y)(\epsilon t)^{n+3}$ had, when y was the catalectic point q_h, an apolar quadratic $(\kappa_h t)^2 = (\kappa_{h1} t)\cdot(\kappa_{h2} t)$. Hence

$$(\beta q_h)(\epsilon t)^{n+3} = k_1(\kappa_{h1} t)^{n+3} + k_2(\kappa_{h2} t)^{n+3}.$$

From this it is at once evident that

$$(3) \qquad (\beta q_h)(\epsilon \kappa_h)^2(\epsilon t)^{n+1} \equiv 0 \qquad \text{(in } t\text{)}.$$

Translating this to variables x', z' in T', we see that

$$(4) \qquad (\alpha' n_h)(\beta' q_h)(\gamma' z') \equiv 0 \qquad \text{(in } z'\text{)},$$

where n_h is the nodular point x' determined by $(\kappa_h t)^2$. Hence

(5) *The trilinear form* $T'(2, n, n + 1)$ *determined by the osculant* $(n + 1)$*-ics of* $r_n^{n+3}(t)$ *has for pairs* x', y, *neutral for* z', *the nodular set* $P'^2_{n+2,2}$ *of* $\rho_2^{n+3}(t)$, *and the catalectic set* $Q^n_{n+2,2}$ *of* $r_n^{n+3}(t)$.

It may be remarked that this theorem does not of itself impose conditions on the nodal set $P^2_{n+2,2}$ of $\rho_2^{n+3}(t)$. It merely states a relation between two sets covariantly related to the nodal set, this relation being sufficient to define the set Q^n_{n+2} in terms of the set $P'^2_{n+2,2}$. Indeed,[5] if the plane $[x']$ is mapped upon a Veronesean $V_2^{(n-1)^2}$ in a space $\left[\binom{n+1}{2} - 1\right]$ by the aggregate of curves of order $n - 1$, the nodular set $P'^2_{n+2,2}$ is mapped upon a set $R_{n+2,2}$ on $V_2^{(n-1)^2}$ whose associated set, uniquely defined to within projectivities, in $[n]$ is $Q^n_{n+2,2}$. The conditions desired arise from the existence of a second trilinear form $T(2, n, n + 1)$ whose neutral pairs are the nodal set $P^2_{n+2,2}$ of $\rho_2^{n+3}(t)$ and the *same* catalectic set $Q^n_{n+2,2}$ of $r_n^{n+3}(t)$.

7. The perspective $(n + 1)$-ics of $\rho_2^{n+3}(t)$ and the second trilinear form $T(2, n, n + 1)$. The first trilinear form T' was obtained in the last section for every n. However, according to §4, the perspective envelopes of $\rho_2^{n+3}(t)$ behave differently according as n is odd or even. We begin therefore with the simpler case of n odd. Then the perspective envelopes of class $\nu = \frac{1}{2}(n + 3)$ and class $\nu = n + 1$ are given respectively by [cf. §4, (3), (6)]

$$(1) \qquad (\pi x)(d\tau)(\delta t)^{\frac{1}{2}(n+3)} = 0,$$

$$(2) \qquad (\pi x)(dm)(\delta t)^{\frac{1}{2}(n+3)}(\mu t)^{\frac{1}{2}(n-1)} = 0,$$

where, in (2), $(m\tau)(\mu t)^{\frac{1}{2}(n-1)}$ is a generic double binary form of the orders indicated.

If x is a point t_1 of $\rho_2^{n+3}(t)$, the factor (tt_1) separates from (1) and (2). If x is a node p_h of $\rho_2^{n+3}(t)$ with parameters $(\kappa_h t)^2$, this factor $(\kappa_h t)^2$ separates from (1), so that

$$(3) \qquad (\pi p_h)(d\tau)(\delta t)^{\frac{1}{2}(n+3)} \equiv (\kappa_h t)^2\cdot(l_h \tau)(\lambda_h t)^{\frac{1}{2}(n-1)}.$$

[5] Cf. Coble, loc. cit. (footnote 2), §5, (6).

In the equation (4) of §4 of $\rho_2^{n+3}(t)$, let ξ be the line $t = t_1$ of the perspective envelope (1). Since this ξ cuts $\rho_2^{n+3}(t)$ in the point $t = t_1$, (tt_1) will factor out after this substitution. We can therefore define a form f of the degrees indicated,

$$(4)\qquad f(\pi^3, \tau^1, t^{n+2}, t_1^{\frac{1}{2}(n+1)}) = (\pi\pi'\pi'')(dd')(\delta t)^{\frac{1}{2}(n+3)}(\delta' t)^{\frac{1}{2}(n+3)}(\delta'' t_1)^{\frac{1}{2}(n+3)}/(tt_1),$$

which for given τ and t_1 furnishes the further $(n + 2)$ collinear points of $\rho_2^{n+3}(t)$ on the line (1) determined by τ and t_1.

A collinear $(n + 2)$-point of $\rho_2^{n+3}(t)$ determines a unique prime η in the space $[n]$ of the conjugate curve. For, the polar of this $(n + 2)$-point with respect to every point section from y of $r_n^{n+3}(t)$ must be apolar to (tt_1), and must therefore either be (tt_1), or vanish identically. Since it must vanish identically for at least ∞^{n-1} points y, there is a prime η whose points y have point sections apolar to the collinear $(n + 2)$-point. On taking the polar of the collinear $(n + 2)$-point in (4) with respect to $(\beta y)(\epsilon t)^{n+3}$ [cf. §5, (1)], eliminating the factor (tt_1), and then replacing t_1 by t, we obtain a form

$$(5)\qquad g(y^1, \tau^1, t^{\frac{1}{2}(n-1)}).$$

Then, from the foregoing, it is clear that

(6) *With τ and t as parameters, $g = 0$ is the parametric equation of a rational surface of class $n - 1$ in $[n]$, the Stahl surface of primes attached to $r_n^{n+3}(t)$. A particular prime is the locus of points y whose $(n + 3)$-ic sections of $r_n^{n+3}(t)$ have a common apolar $(n + 2)$-ic, this being the $n + 2$ parameters of points cut out on $\rho_2^{n+3}(t)$ by the line τ, t in* (1) *in addition to t itself.*

This Stahl envelope is the surface S of §3 and we apply it as there indicated, to set up a coördinate system in $[n]$ in such a way that a form $(m\tau)(\mu t)^{\frac{1}{2}(n-1)}$ is the locus of points $y = m_{i,j}$ on the prime τ, t of the envelope. If then in the equation (2) of the system of perspective $(n + 1)$-ics of $\rho_2^{n+3}(t)$ we replace the coefficients $m_{i,j}$ of $(m\tau)(\mu t)^{\frac{1}{2}(n-1)}$ by coördinates y in $[n]$ referred to the Stahl surface, and if further, as in (6) of §2, we replace the $(n + 1)$-ic combinations of t by coördinates z in a space $[n + 1]$ referred to a norm-curve N^{n+1}, then the equation (2) becomes a trilinear form

$$(7)\qquad T(2, n, n + 1) = (\alpha x)(\beta y)(\gamma z) = 0.$$

We seek the neutral pairs x, y of this trilinear form T. We observe that $g = 0$ determines, for given y, the ∞^1 pairs τ, t each of which defines in (6) a line which cuts $\rho_2^{n+3}(t)$ in the point t and a further collinear $(n + 2)$-ic which is apolar to the $(n + 3)$-ic section of $r_n^{n+3}(t)$ by the given y. If y is a catalectic point q_h, this section is $(\beta q_h)(\epsilon t)^{n+3} = k_1(\kappa_{h1} t)^{n+3} + k_2(\kappa_{h2} t)^{n+3}$ (cf. §6). Every $(n + 2)$-ic apolar to such an $(n + 3)$-ic must contain the factor $(\kappa_h t)^2$, and the collinear $(n + 2)$-ic must be on the node p_h. But, according to (3), the line (1) is on the node p_h if $(l_h\tau)(\lambda_h t)^{\frac{1}{2}(n-1)} = 0$. Hence

(8) *The catalectic point q_h of $r_n^{n+3}(t)$, when referred to the coördinate system defined as in* §3 *by the Stahl envelope* (5), *is given by the coefficients of the form $(l_h\tau)(\lambda_h t)^{\frac{1}{2}(n-1)}$ in* (3).

Now, the form (2), which is interpreted in (7) as the trilinear form T, is merely the bilinear invariant with respect to τ of $(\pi x)(d\tau)(\delta t)^{\frac{1}{2}(n+3)}$ and $(m\tau)(\mu t)^{\frac{1}{2}(n-1)}$. When x is at p_h, and $y = m_{i,j}$ is at q_h, these reduce respectively to $(\kappa_h t)^2\cdot(l_h\tau)\cdot(\lambda_h t)^{\frac{1}{2}(n-1)}$ and $(l_h\tau)(\lambda_h t)^{\frac{1}{2}(n-1)}$. Hence their bilinear invariant with respect to τ vanishes identically in t. Hence also $(\alpha p_h)(\beta q_h)(\gamma z)$ vanishes identically in z. Thus we have proved the theorem complementary to (5) of §6:

(9) *The trilinear form $T(2, n, n + 1)$ determined by the perspective $(n + 1)$-ics of $\rho_2^{n+3}(t)$ has for pairs x, y, neutral for z, the nodal set $P_{n+2,2}^2$ of $\rho_2^{n+3}(t)$ and the catalectic set $Q_{n+2,2}^n$ of $r_n^{n+3}(t)$.*

The modifications in the above proof of (9) for the case when n is odd, which are necessary for the case when n is even, are in great measure merely formal, and of the sort already used in (3′), (3′.1) of §4, yet it seems necessary to carry them through up to a certain point. Thus the systems of perspective curves in (1) and (2) are replaced as in §4 by

$$(1') \qquad (\pi x)(d\tau)(\delta t)^{\frac{1}{2}(n+4)} = 0,$$

$$(1'.1) \qquad (\pi x)(d\tau^{(0)})(\delta t^{(0)})^{\frac{1}{2}(n+4)} \equiv 0;$$

$$(2') \qquad (\pi x)(dm)(\delta t)^{\frac{1}{2}(n+4)}(\mu t)^{\frac{1}{2}n}/(tt^{(0)}) = 0,$$

$$(2'.1) \qquad (m\tau^{(0)})(\mu t^{(0)})^{\frac{1}{2}n} = 0.$$

We have again in (2′) the ∞^n perspective envelopes of class $n + 1$ determined by the ∞^n forms with coefficients $m_{i,j}$ conditioned as in (2′.1).

When, in (1′), (1′.1), x is at the node p_h, we find that

$$(3') \qquad (\pi p_h)(d\tau)(\delta t)^{\frac{1}{2}(n+4)} = (\kappa_h t)^2\cdot(l_h\tau)(\lambda_h t)^{\frac{1}{2}n},$$

$$(3'.1) \qquad (l_h\tau^{(0)})(\lambda_h t^{(0)})^{\frac{1}{2}n} = 0.$$

When in (4′) of §4 we replace ξ by the line $(\pi x)(d\tau)(\delta t_1)^{\frac{1}{2}(n+4)}$, we get, after dropping the factor (tt_1), a form

$$(4') \qquad f(\pi^3, \tau^1, t^{n+2}, t_1^{\frac{1}{2}(n+2)}),$$

$$(4'.1) \qquad f(\pi^3, \tau^{(0)1}, t^{n+2}, t_1^{(0)\frac{1}{2}(n+2)}) \equiv 0 \qquad \text{(in } t\text{)},$$

since the line has a factor $(t_1t^{(0)})$ when $\tau = \tau^{(0)}$. When $f = 0$ in (4′), we have, for given t_1 and variable τ, the pencil of collinear $(n + 2)$-ics cut out on $\rho_2^{n+3}(t)$ by lines on the given point t_1 of $\rho_2^{n+3}(t)$.

The Stahl envelope, ∞^2 primes in one-one correspondence with the collinear $(n + 2)$-points of $\rho_2^{n+3}(t)$, is obtained as before by taking the polar of (4′) with respect to the conjugate envelope $(\beta y)(\epsilon t)^{n+3}$, factoring out (tt_1), and replacing t_1 by t, to get

$$(5') \qquad g(y^1, \tau^1, t^{\frac{1}{2}n}),$$

$$(5'.1) \qquad g(y^1, \tau^{(0)1}, t^{(0)\frac{1}{2}n}) \equiv 0 \qquad \text{(in } y\text{)},$$

this subsidiary condition arising from (4′.1).

Then the theorem (6) holds as it is stated. The Stahl envelope is not, however, of the sort used before with $\frac{1}{2}(n-1)$ replaced by $\frac{1}{2}n$. Such a surface would be of class n in $[n+1]$. It is rather the section of such a surface by one of its primes and, indeed, by precisely the prime $\tau^{(0)}, t^{(0)}$. The coördinate system $y = m_{i,j}$ which applies to this case is that described in (8) of §3. With such a coördinate system we again get the trilinear form $T(2, n, n+1)$ in (7), and carry out the proofs of (8) and (9) as before. Hence

(10) *With* (6), (7), (8) *modified as indicated above, the theorem* (9) *is valid for both even and odd values of* n.

8. **Conditions on the nodal set $P^2_{n+2,2}$ of $\rho_2^{n+3}(t)$.** We are now in position to state actual conditions on the nodal set $P^2_{n+2,2}$ of $\rho_2^{n+3}(t)$. Both trilinear forms T, T' have in the $[n]$ of y the *same* neutral set, the catalectic set $Q^n_{n+2,2}$. This is simply associated with, and therefore projectively defines, a set $R_{n+2,2}$ in a space $\left[\binom{n+1}{2}-1\right]$. Because of the existence of T, the plane of $\rho_2^{n+3}(t)$ can be mapped by curves of order $n-1$ upon a Veronesean $V_2^{(n-1)^2}$ in $\left[\binom{n+1}{2}-1\right]$ in such a way that the nodal set $P^2_{n+2,2}$ maps into the set $R_{n+2,2}$ on $V_2^{(n-1)^2}$ [cf. §6, (5), et seq.]. Because of the existence of T', the plane of the nodular set can be similarly mapped into a Veronesean $V_2'^{(n-1)^2}$, the nodular set $P'^2_{n+2,2}$ mapping into the same set $R_{n+2,2}$. Hence the conditions sought are expressed as follows:

(1) *If the plane of $\rho_2^{n+3}(t)$ is mapped by curves of order $n-1$ upon the Veronesean $V_2^{(n-1)^2}$ in $\left[\binom{n+1}{2}-1\right]$, the nodal set P mapping into a set R, then on this set R there is a second Veronesean $V_2'^{(n-1)^2}$, R on $V_2'^{(n-1)^2}$ being the map of the nodular set P'.*

The symmetry of this condition with respect to the nodal set P and the nodular set P' suggests that the nodular set P' in $[x']$ is also the nodal set P of a rational curve $\rho_2'^{n+3}(t)$ in $[x']$. This indeed is true for the rational sextic $(n = 3)$. For, in this case, each V_2^4 is the double surface of a cubic primal M_4^3 in [5], and each M_4^3 cuts the double V_2^4 of the other M_4^3 in a curve of order 12 with ten nodes at R^5_{10}. Thus we have the maps of rational sextics with nodes at P^2_{10} in $[x]$, and at P'^2_{10} in $[x']$. This situation has been obtained in another paper[6] from an entirely different point of view.

Also in the case of a rational sextic the conditions given in (1) are independent. There are three conditions on R^5_{10} that it be on V_2^4, and three more that it be on $V_2'^4$. Thus each of R^5_{10}, and its associated Q^3_{10}, the ten nodes of a symmetroid, is subject to six conditions, whereas P^2_{10} is subject only to the three conditions that R^2_{10} on V_2^4 is also on $V_2'^4$, and P'^2_{10} is subject to the three additional condi-

[6] A. B. Coble, *Associated sets of points*, Trans. Amer. Math. Soc., vol. 24(1922), pp. 1-20. See Theorem 26.

tions. For higher values of n the existence of the second $V_2^{(n-1)^2}$ on $R_{n+2,2}$ no longer imposes independent conditions on $R_{n+2,2}$. For, $V_2^{(n-1)^2}$ admitting a collineation g_8 depends upon $\binom{n+1}{2}^2 - 9$ constants, and there are $\binom{n+1}{2} - 3$ conditions that $V_2^{(n-1)^2}$ be on a point. Thus there are a finite number of $V_2^{(n-1)^2}$'s on $\binom{n+1}{2} + 3$ points, and $(n-2)\left[\binom{n+1}{2} - 3\right] = \frac{1}{2}(n-2)^2(n+3)$ conditions that the remaining $n-2$ points of $R_{n+2,2}$ be on one of these $V_2^{(n-1)^2}$'s. But this is precisely the number of conditions on the set $Q_{n+2,2}^n$ of the neutral sets $P_{n+2,2}^2$, $Q_{n+2,2}^n$ of a form $T(2, n, n+1)$, this form being defined by the set $P_{n+2,2}^2$ alone.[7] But the $(n-2)\left[\binom{n+1}{2} - 3\right]$ conditions that $R_{n+2,2}$ be on $V_2'^{(n-1)^2}$ cannot be altogether independent of the conditions that R be on V if $n > 3$, since then these conditions would fall on $P_{n+2,2}^2$ and reduce the absolute constants in $P_{n+2,2}^2$ to a smaller number than the $3n$ which a nodal set $P_{n+2,2}^2$ actually possesses.

However, it does not seem likely that the conditions (1) are sufficient, for the reason that, beyond $n = 3$, there seems to be no evidence that rational curves are paired as the condition suggests.

9. **The transformation of coördinates defined by the two Veroneseans on** $R_{n+2,2}$. In the mapping of $[x]$ upon $V_2^{(n-1)^2}$ in $\left[\binom{n+1}{2} - 1\right]$ we have a natural coördinate system in $\left[\binom{n+1}{2} - 1\right]$ set up by the mapping

(1) $$X_{ijk} = x_0^i x_1^j x_2^k \qquad (i + j + k = n).$$

This yields a point X on $V_2^{(n-1)^2}$ for each point x in the plane of $P_{n+2,2}^2$. However, these combinations of x are merely the coefficients of a form $(x\xi)^{n-1}$ of class $n-1$ so that the mapping implies a coördinate system of the following type:

(2) *If* $(a\xi)^{n-1}$ *and* $(\alpha x)^{n-1}$ *are ternary forms of class and order* $n-1$, *then*

$$X_{ijk} = a_0^i a_1^j a_2^k, \qquad U_{ijk} = \binom{n-1}{i, j, k} \alpha_0^i \alpha_1^j \alpha_2^k$$

is a mapping of these forms upon the points and primes of $\left[\binom{n+1}{2} - 1\right]$ *with the incidence condition*

$$(XU) = (a\alpha)^{n-1}.$$

[7] Cf. Coble, loc. cit. (footnote 2), §5, (6).

In these coördinates X, U a change of coördinates is represented by a generic form

$$(3) \qquad C = (\gamma x)^{n-1}(c\xi)^{n-1}.$$

For, given any point X, an $(n-1)$-ic, $(a\xi)^{n-1}$, the form defines a new $(n-1)$-ic, $(\gamma a)^{n-1}(c\xi)^{n-1}$, whose coefficients X' are generic linear expressions in the coefficients of $(a\xi)^{n-1}$.

We had above in $\left[\binom{n+1}{2}-1\right]$ a second surface $V_2'^{(n-1)^2}$ which equally well sets up a coördinate system X', U' in $\left[\binom{n+1}{2}-1\right]$. Thus any point in $\left[\binom{n+1}{2}-1\right]$ has coördinates X, X' when referred to V, V' respectively. We seek the algebraic expression for the transformation of coördinates from X to X', i.e., a form

$$(4) \qquad C = (\gamma x)^{n-1}(c'\xi')^{n-1}.$$

Since $x = p_h$ and $x' = n_h$ map into the *same* point $X = X'$ of V, V', it is sufficient to require that C be such that

$$(5) \qquad (\gamma p_h)(c'\xi')^{n-1} \equiv (n_h\xi')^{n-1} \quad \left(h = 1, \cdots, \binom{n+2}{2}\right).$$

For, the change of coördinates is defined by merely $\binom{n+1}{2}+1 < \binom{n+2}{2}$ corresponding pairs.

We begin with the trilinear form T of (9) of §7 which arises from the perspective $(n+1)$-ic envelopes of $\rho_2^{n+3}(t)$ in (2), (2′) of §7. In this let x be fixed and $y = m_{i,j}$ in $[n]$ be variable. For variable t, the prime in $[n]$ on the Stahl envelope runs over an $(n+1)$-ic rational envelope in $[n]$. The "conjugate" form of this envelope is a single form $H(x^{n+1}, t^{n+1})$, whose polar form

$$(6) \qquad H(x^{n+1}; t_1, t_2, \cdots, t_{n+1}),$$

equated to zero, is the condition that the primes $t_1, \cdots, t_{n+1}$ in $[n]$ are on a point y. If x is on the point t_1 of $\rho_2^{n+3}(t)$, then (tt_1) factors out of T in (2), (2′) of §7, and $H(x^{n+1}, t^{n+1}) = (tt_1)^{n+1}$ and the polar of t_1 vanish whatever $t_2, \cdots, t_{n+1}$ are. If x is the node p_h of $\rho_2^{n+3}(t)$, then $(\kappa_h t)^2$ factors out of T, whence $H(p_h^{n+1}, t^{n+1}) \equiv 0$ in t. Hence

(7) *For variable x, $H = 0$ in* (6) *is the adjoint $(n+1)$-ic curve on the points $t_1, t_2, \cdots, t_{n+1}$ of $\rho_2^{n+3}(t)$.*

We write H in the form

$$(8) \qquad H(x^{n+1}, t^{n+1}) = H_0 t^{n+1} + \binom{n+1}{1} H_1 t^n + \binom{n+2}{2} H_2 t^{n-1} + \cdots + H_n.$$

Since H is a perfect $(n+1)$-th power when x is on

$$(9) \qquad \rho_2^{n+3}(t) = R^{n+3}(x),$$

there follows that

(10) *The two-row determinants of the matrix*

$$\begin{vmatrix} H_0 & H_1 & H_2 & \cdots & H_n \\ H_1 & H_2 & H_3 & \cdots & H_{n+1} \end{vmatrix}$$

all contain $R^{n+3}(x)$ as a factor with residual factors of order $n-1$.

We consider now the quadratic system (parameter t)

$$(11) \qquad H(x^{n+1}; t_1, \cdots, t_{n-1}, t^2) = 0$$

of adjoints on the fixed points $t_1, t_2, \cdots, t_{n-1}$ of $\rho_2^{n+3}(t)$. The envelope of this system is the discriminant of the quadratic in t, and, after separation of the factor $R^{n+3}(x)$, this has the form

$$(12) \qquad C(x^{n-1}; t_1^2, t_2^2, \cdots, t_{n-1}^2) = 0.$$

An expression for this form C in terms of the coefficients H of the binary form (8) is

$$(12') \quad C(x^{n-1}; t_1^2, \cdots, t_{n-1}^2) = (HH')^2(Ht_1) \cdots (Ht_{n-1})(H't_1) \cdots (H't_{n-1})/R^{n+3}(x).$$

Another form of (12) is found by observing that it is the condition that $H(x^{n+1}, t^{n+1})$ have an apolar n-ic $(\alpha t)^n$ of which $(tt_1), \cdots, (tt_{n-1})$ are factors. On expressing these $n+1$ conditions, and eliminating the $n+1$ coefficients of $(\alpha t)^n$, we get C in the determinant form:

$$C(x^{n-1}; t_1^2, \cdots, t_{n-1}^2)$$

$$(12'') \qquad = \begin{vmatrix} H_0 & H_1 & H_2 & \cdots & H_n \\ H_1 & H_2 & H_3 & \cdots & H_{n+1} \\ 1 & -t_1 & t_1^2 & \cdots & (-1)^n t_1^n \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ 1 & -t_{n-1} & t_{n-1}^2 & \cdots & (-1)^n t_{n-1}^n \end{vmatrix} / R^{n+3}(x) \cdot \prod (t_i t_j).$$

If, in (6), we let t_{n+1} be variable, we have a pencil of adjoint $(n+1)$-ic curves on the $\binom{n+2}{2}$ nodes and the n points $t_1, \cdots, t_n$ of $\rho_2^{n+3}(t)$, and on $\binom{n}{2}$ further base points x. We shall call such a pencil $\pi(t_1, \cdots, t_n)$. For each x, and $t_{n+1} = t = t_n$, equation (11) is satisfied. Hence

(13) *For given x, $C = 0$ in (12) is the equation of the I_{n-2}^n of pencils $\pi(t_1, \cdots, t_n)$ which have a base point at x. For given $t_1, \cdots, t_{n-1}$, it is the $(n-1)$-ic locus of $\binom{n}{2}$ further base points of pencils $\pi(t_1, \cdots, t_{n-1}, t)$.*

We consider the pencils π which have a node at p_h, a node of $\rho_2^{n+3}(t)$. For this it is necessary and sufficient that the two roots $(\kappa_{h1}t)$, $(\kappa_{h2}t)$ of $(\kappa_h t)^2$ be included in the $n + 1$ further intersections of the adjoint (6). This is just two conditions, but a pencil of such adjoints will have a fourth intersection at the node, so that one of the $\binom{n}{2}$ further base points x of the pencil has moved up to p_h. Thus the pencils $\pi(\kappa_{h1}, \kappa_{h2}, t_3, \cdots, t_n)$ have one of their $\binom{n}{2}$ further base points at $x = p_h$, where $t_3, \cdots, t_n$ are arbitrary. Hence the I_{n-2}^n of (13) has a neutral pair $(\kappa_h t)^2$, i.e., the form (12) for $x = p_h$ becomes

$$(14) \qquad C(p_h^{n-1}; t_1^2, \cdots, t_{n-1}^2) = (\kappa_h t_1)^2\cdot(\kappa_h t_2)^2\cdot \cdots \cdot(\kappa_h t_{n-1})^2.$$

We now interpret the involution form $C(x^{n-1}; t_1^2, \cdots, t_{n-1}^2)$ as a polarized envelope in the [2] of x' referred to the norm-conic $K(t)$. A generic envelope $(a'\xi')^{n-1}$ can be polarized to yield $(a'\xi')(a'\xi'') \cdots (a'\xi^{(n-1)})$. In this let ξ', ξ'', $\cdots$, $\xi^{(n-1)}$ be lines $\xi' = t_1$, $\xi'' = t_2$, $\cdots$, $\xi^{(n-1)} = t_{n-1}$ of $K(t)$. Then the polarized envelope becomes a symmetric form of degree two in each of $t_1, \cdots, t_{n-1}$, and any such symmetric form can be interpreted in this way as a polarized envelope. This envelope is the *apolarity primal* defined in §2. Thus, for every x we have an $(n - 1)$-ic envelope in the [2] of x', and therefore a form C of the kind desired in (4). In this notation, $(\kappa_h t_1)^2 = (n_h\xi')$, $(\kappa_h t_2)^2 = (n_h\xi'')$, etc. Hence

(15) *If the form C in* (12), *as a symmetric form in $t_1, \cdots, t_{n-1}$, is interpreted as the apolarity locus with respect to $K(t)$ of an $(n - 1)$-ic envelope, then C becomes a form*

$$C = (\gamma x)^{n-1}(c'\xi')^{n-1},$$

with the subsidiary property that $(\gamma p_h)^{n-1}(c'\xi')^{n-1} = (n_h\xi')^{n-1}$. As noted above C determines the transformation of coördinates in the $\left[\binom{n+1}{2} - 1\right]$ of $R_{n+2,2}$ induced by the change from the one to the other of the two Veroneseans on $R_{n+2,2}$.

We have introduced this transformation, whose existence is a necessary consequence of the two Veroneseans on $R_{n+2,2}$, as a check on the entire exposition above. It is clear that a form $C = (\gamma x)^{n-1}(c'\xi')^{n-1}$, which becomes a perfect power $(n_h\xi')^{n-1}$ for $\binom{n+2}{2}$ points $x = p_h$, is highly restricted. That it could be obtained from such relatively simple considerations was rather unexpected.

University of Illinois.

THE METHOD OF ORTHOGONAL PROJECTION IN POTENTIAL THEORY

By Hermann Weyl

1. **Stating the problem.** Roughly speaking the classical boundary-value problem of potential theory for a region G in the Cartesian (x_1, x_2, x_3)-space consists in splitting a given function φ in G into two summands $\psi + \eta$, the first of which vanishes along the boundary of G, while the second η is harmonic. The two components are orthogonal if a metric in the functional space is based upon the Dirichlet integral

$$D[\varphi] = \int (\text{grad } \varphi)^2.$$

$\int$ indicates integration over G.

This fact suggests the idea of replacing the scalar φ by the vector field

$$f = \text{grad } \varphi \qquad \left(\text{in components: } f_i = \frac{\partial \varphi}{\partial x_i}\right)$$

and of operating in the Hilbert space of all vector fields f, with its metric defined by

$$\|f\|^2 = \int f^2 = \int (f_1^2 + f_2^2 + f_3^2).$$

Then the question arises how to characterize a vector field f as a gradient field without assuming more than its Lebesgue integrability. The vanishing of the line integral

$$\int (f \cdot dx) = \int (f_1 \, dx_1 + f_2 \, dx_2 + f_3 \, dx_3)$$

over any closed curve in G will not do, because we have nothing but *spatial* integration at our disposal. The customary condition

$$\text{rot } f = 0 \tag{1}$$

uses differentiation. Let v be any vector field vanishing at the boundary of G. The formula

$$\text{div } [f, v] = (v \cdot \text{rot } f) - (f \cdot \text{rot } v) \tag{2}$$

Received July 13, 1940.

for the vector product $[f, v]$ with its integrated consequence

$$\int (v \cdot \operatorname{rot} f) = \int (f \cdot \operatorname{rot} v)$$

shows that (1) is equivalent to the relation

$$\int (f \cdot \operatorname{rot} v) = 0$$

holding for all fields v of the above-described nature. This characterization satisfies our demands. While the vector field f itself is single-valued, the potential φ whose existence is assured by (1) may be multivalued with certain "periods", if G is not simply connected. However, this is to the good because it makes our problem more inclusive.

After these preliminaries we are now going to fix our subject in precise terms. G is any open set in 3-space. All functions which we consider are understood to be defined and their squares to be Lebesgue integrable in the interior of G; and equality of functions is understood as prevailing "almost everywhere", i.e., except in a set of Lebesgue measure zero. A continuous function ψ with continuous (first) derivatives which vanishes in a boundary strip, i.e., outside some compact subset G^* of G, is said to be of class Γ, and so is a vector field whose components are of class Γ. A vector field f is called *irrotational* or *solenoidal* if it satisfies the condition

$$\int (f \cdot \operatorname{rot} v) = 0 \quad \text{for every vector field } v \text{ of class } \Gamma \tag{3}$$

or

$$\int (f \cdot \operatorname{grad} \psi) = 0 \quad \text{for every scalar field } \psi \text{ of class } \Gamma \tag{4}$$

respectively. We state the almost trivial

LEMMA 1.

$$\int (\operatorname{grad} \psi \cdot \operatorname{rot} v) = 0 \tag{5}$$

for any scalar and vector fields ψ, v of class Γ.

Our pivotal proposition is as follows:

THEOREM I. *A field which is both irrotational and solenoidal equals a field f possessing derivatives of all orders and satisfying the equations*

$$\operatorname{div} f = 0, \qquad \operatorname{rot} f = 0. \tag{6}$$

Because of the equation

$$\Delta f = \operatorname{grad} \operatorname{div} f - \operatorname{rot} \operatorname{rot} f \tag{7}$$

for the Laplace operator Δ working on the components of f, these components are harmonic functions.

Let $\mathfrak{F}_0$ be the complete Hilbert space of all vector fields f in G of finite $\|f\|$. The elements of $\mathfrak{F}_0$ which are irrotational or solenoidal, or both, form complete subspaces $\mathfrak{F}$, $\mathfrak{F}'$, $\mathfrak{E}$ of $\mathfrak{F}_0$. Our Theorem I describes the elements of $\mathfrak{E}$. The closure (in the sense of the metric $\|f\|^2$) of the fields

$$\text{(8)} \qquad \text{grad } \psi \qquad (\psi \text{ any scalar of class } \Gamma)$$

will be denoted by $\mathfrak{G}$, the closure of the fields

$$\text{(9)} \qquad \text{rot } v \qquad (v \text{ any vector of class } \Gamma)$$

by $\mathfrak{G}'$. According to Lemma 1, $\mathfrak{G}$ is part of $\mathfrak{F}$, and $\mathfrak{G}'$ part of $\mathfrak{F}'$. By the very definitions (3), (4) we have the following decompositions into mutually orthogonal components:

THEOREM II.

$$\mathfrak{F} = \mathfrak{G} + \mathfrak{E}, \qquad \mathfrak{F}' = \mathfrak{G}' + \mathfrak{E};$$

$$\mathfrak{F}_0 = \mathfrak{G}' + \mathfrak{F}, \qquad \mathfrak{F}_0 = \mathfrak{G} + \mathfrak{F}'.$$

The first equation contains essentially the solution of the boundary-value problem in our generalized form. The second equation is no less important. We shall treat both problems along parallel lines.

In (8) the potential ψ of class Γ is uniquely determined by the field $f = \text{grad } \psi$. Not so for (9); here one has to add div v to rot v in order to fix v. Hence we are led to consider pairs $\mathfrak{f} = (f, \varphi)$ consisting of any solenoidal vector f and any scalar φ, with the metric of the complete Hilbert space $\mathfrak{F}^+$ of pairs $\mathfrak{f}$ defined by

$$\|\mathfrak{f}\|^2 = \int f^2 + \int \varphi^2.$$

Let $\mathfrak{G}^+$ denote the closure (in the sense of this metric) of the pairs of the following form

$$f = \text{rot } v, \qquad \varphi = \text{div } v \qquad (v \text{ any field of class } \Gamma),$$

while an element (f, φ) of $\mathfrak{F}^+$ is said to lie in $\mathfrak{E}^+$ if it satisfies the equation

$$\text{(10)} \qquad \int (f \cdot \text{rot } v) + \int \varphi \cdot \text{div } v = 0$$

for any v of class Γ. We have the decomposition

$$\text{(III)} \qquad \mathfrak{F}^+ = \mathfrak{G}^+ + \mathfrak{E}^+$$

into two mutually orthogonal constituents. Theorem I is paralleled by the following statement about $\mathfrak{E}^+$:

THEOREM III. *Any element of $\mathfrak{E}^+$ equals a pair (f, φ) whose members f and φ have derivatives of all orders and are linked by the equations*

$$\text{div}\, f = 0, \qquad \text{rot}\, f = \text{grad}\, \varphi. \tag{11}$$

These equations show at once that φ and the components of f are harmonic. We speak of the splitting (III) as the solution of the third boundary-value problem.

The proofs of our three theorems follow in §2. Lemma 1 will be settled in §3 in connection with a general survey of vector analysis. Some important auxiliary inequalities, the first of which is due to H. Poincaré, are discussed in §4. After these preliminaries a new question, that of topological periods, first formulated at the end of §2, is taken up in §§5 and 6. §7 deals with the behavior at the boundary, and §8 contains concluding remarks about 2 and n dimensions.

The idea of replacing Dirichlet's minimum principle by the construction of orthogonal projection in a suitable Hilbert space seems to have occurred first to O. Nikodym,[1] who applied it to a simpler problem than ours. Chevalley tells me that he and de Possel some years ago developed potential theory along similar lines without publishing their investigations. I depend, above all, on two papers by K. Friedrichs.[2] In particular, the proof of Lemma 2 is a modification of a construction due to Fubini, Courant and Friedrichs. In his second paper Friedrichs overcomes the difficulties involved in the differentiation connecting grad φ with φ by some ingeniously devised "mollifiers" J_a. Our way of dealing directly with the space of vector fields f and expressing their irrotational character by (3) removes an undesirable limitation and opens the road for a parallel treatment of our two boundary-value problems. By the question of periods our investigation is linked to Hodge's construction of harmonic differential forms with preassigned periods.[3]

The two constructions which serve to prove Lemma 2 and Theorem VII form the backbone of this paper.

2. Proof of the central theorems.

Suppose f is both irrotational and solenoidal. Let w be a vector field of class Γ_2, i.e., vanishing in a boundary strip and continuous with its derivatives up to the second order. Replace ψ by div w in (4) and v by rot w in (3) and subtract; because of (7),

$$\Delta w = \text{grad div}\ w - \text{rot rot}\ w,$$

[1] *Sur un théorème de M. S. Zaremba concernant les fonctions harmoniques*, Journal de Mathématiques, (9), vol. 12(1933), pp. 95–109.

[2] *On certain inequalities and characteristic value problems for analytic functions and for functions of two variables*, Trans. Amer. Math. Soc., vol. 41(1937), pp. 321–364; *On differential operators in Hilbert spaces*, Amer. Jour. Math., vol. 61(1939), pp. 523–544.

[3] W. V. D. Hodge, *A Dirichlet problem for harmonic functionals, with applications to analytic varieties*, Proc. London Math. Soc., (2), vol. 36(1933), pp. 257–303.

one gets the relation

$$\int (f \cdot \Delta w) = 0$$

in which the three components separate, and finds oneself called upon to prove this

LEMMA 2. *A scalar η satisfying the equation*

$$\int \eta \cdot \Delta \zeta = 0 \tag{12}$$

for every scalar ζ of class Γ_2 equals a harmonic function.

This once accomplished, we conclude that the components of f are harmonic functions with derivatives of all orders, and then, owing to (2) and

$$\operatorname{div}(\psi f) = (f \cdot \operatorname{grad} \psi) + \psi \cdot \operatorname{div} f, \tag{13}$$

the equations (4), (3) lead back to

$$\operatorname{div} f = 0 \quad \text{and} \quad \operatorname{rot} f = 0.$$

The proof of Lemma 2 depends on the skillful construction of scalars ζ of class Γ_2. Consider a sphere K of radius R in G whose center we take as the origin, and let x' be a point within the "cavity" K. Green's function $G(x, x'; R)$ representing the potential in K of a point-charge at x' is the difference of two parts, the singular part

$$\frac{1}{|x - x'|} \tag{14}$$

and the compensating part

$$L(x, x'; R) = \begin{cases} \dfrac{R}{(R^4 - 2R^2(x \cdot x') + x^2 \cdot x'^2)^{\frac12}} & \text{for } |x| \leqq R, \\[2ex] \dfrac{1}{(x^2 - 2(x \cdot x') + x'^2)^{\frac12}} & \text{for } |x| \geqq R. \end{cases}$$

Here x denotes the vector from the origin to the point x. A surface density

$$-\frac{1}{R} \cdot \frac{R^2 - x'^2}{|x - x'|^3}$$

screens the exterior of K. Replacing the metal surface K by a wall of some thickness, and thus letting R vary between limits a and $b > a$, we form

$$L^*(x, x') = \int_a^b L(x, x'; R)\, dR \Big/ \int_a^b dR. \tag{15}$$

x' is supposed to lie in the cavity $|x| < a$. A spatial density

$$-\frac{1}{b-a}\cdot\frac{x^2 - x'^2}{|x - x'|^3}\cdot\frac{1}{|x|}$$

distributed in the wall $a \leqq |x| \leqq |b|$ now screens the exterior $|x| > b$. Turning to the singular part (13), we replace the point-source x' by a little spherical conductor of radius ρ around x'. Then the following function

$$g(x;\rho) = \begin{cases} \dfrac{1}{|x - x'|} & \text{for } |x - x'| \geqq \rho, \\ \dfrac{1}{\rho} & \text{for } |x - x'| \leqq \rho \end{cases}$$

with the uniform surface density ρ^{-2} on the boundary, takes the place of (14). Again we want spatial, not surface, distribution of sources: charges uniformly distributed over a solid sphere of radius c around x' generate the potential

$$g^*(x) = \int_0^c g(x;\rho)\,\rho^2\,d\rho \Big/ \int_0^c \rho^2\,d\rho, \tag{16}$$

more explicitly,

$$g^*(x, x'; c) = \begin{cases} \dfrac{3}{2}\dfrac{1}{c} - \dfrac{1}{2}\dfrac{|x - x'|^2}{c^3} & \text{for } |x - x'| \leqq c, \\ \dfrac{1}{|x - x'|} & \text{for } |x - x'| \geqq c. \end{cases}$$

We choose

$$\zeta(x) = g^* - L^* \tag{17}$$

in (12) and then find

$$\mathfrak{M}\eta(x) = \frac{1}{4\pi(b-a)}\cdot\int_x \frac{x^2 - x'^2}{|x - x'|^3}\cdot\frac{\eta(x)}{|x|}. \tag{18}$$

$\mathfrak{M}$ is the mean value over the sphere $|x - x'| \leqq c$, while the integral at the right extends over the wall. The right side is a harmonic function $\eta^*(x')$ of x' and independent of c (while the left side is independent of a and b). Since the mean value of the harmonic function $\eta^*(x)$ over a sphere around x' equals $\eta^*(x')$, we find that the integrals

$$\int (\eta(x) - \eta^*(x))$$

extending over any solid spheres in G vanish, and hence $\eta(x)$ and $\eta^*(x)$ coincide almost everywhere, or $\eta(x)$ equals a harmonic function throughout G.

The choice of the differential dR in the integration (15) with respect to R is somewhat arbitrary: it could have been replaced by $\mu(R)\cdot R\,dR$ with any

positive continuous factor $\mu(R)$. The same applies to (16). Instead of (18) one would get

$$\frac{\int \lambda(|x - x'|)\cdot\eta(x)}{\int \lambda(\rho)\,\rho^2\,d\rho} = \int \frac{x^2 - x'^2}{|x - x'|^3}\mu(|x|)\eta(x) \Big/ \int \mu(R)\,R\,dR. \tag{19}$$

$\lambda(\rho)$ and $\mu(R)$ are functions in the intervals $0 \leqq \rho \leqq c$ and $a \leqq R \leqq b$ respectively. Yet the result is not essentially more general than (18) in which one is allowed to vary c as well as a and b.

The function (17) is not exactly of class Γ_2. However, one can extend the class Γ in (3) and (4) to cover also continuous functions which have only piecewise continuous derivatives, with discontinuities on such regular surfaces as spheres. Or sticking to the class Γ one can use the formula (19) with such continuous functions $\lambda(\rho)$ and $\mu(R)$ which vanish at the ends c and a, b of their respective intervals of definition.

Only little modification is required for the proof of Theorem III. The same substitutions $\psi = \operatorname{div} w$ and $v = \operatorname{rot} w$ take place in (4) and (10) respectively, resulting in

$$\int (f\cdot\Delta w) = 0$$

while the substitution $v = \operatorname{grad} \zeta$ (ζ of class Γ_2) in (10) yields

$$\int \varphi\cdot\Delta\zeta = 0.$$

Hence Lemma 2 shows that φ and the components of f are harmonic. Making use of their derivatives we may change the relation (4) holding for all ψ of class Γ into

$$\int \psi\cdot\operatorname{div} f = 0 \quad \text{or} \quad \operatorname{div} f = 0$$

and (10) into

$$\int v(\operatorname{rot} f - \operatorname{grad} \varphi) = 0 \quad \text{or} \quad \operatorname{rot} f - \operatorname{grad} \varphi = 0.$$

We add some elementary observations concerning the local structure of harmonic fields f which satisfy the equations (6).

Take a cube T in G the center of which is chosen as the origin. On account of $\operatorname{rot} f = 0$ the field f is the gradient of a scalar η in T which vanishes in the origin. $\eta(x)$ may be constructed as the line integral of f along the radius (r) joining x with the origin,

$$\eta = \int_0^r f_r\,dr,$$

or more explicitly,

$$\eta(x) = \int_0^1 F(x; \tau)\, d\tau \tag{20}$$

after setting

$$\sum_i x_i \cdot f_i(\tau x_1, \tau x_2, \tau x_3) = F(x; \tau).$$

One easily verifies the equations

$$f_i = \frac{\partial \eta}{\partial x_i}$$

by utilizing the condition rot $f = 0$. Because of div $f = 0$ one has $\Delta\eta = 0$; consequently η itself is harmonic. Moreover, for any harmonic function η in T, the radial field

$$j = x \cdot \eta^* \qquad \text{where } \eta^* = \frac{1}{r}\int_0^r \eta\, dr$$

is, as one readily verifies, a solution of the equation

$$\text{rot rot } j = \text{grad } \eta.$$

Again, in more explicit form,

$$\eta^*(x) = \int_0^1 \eta(\tau x_1, \tau x_2, \tau x_3)\, d\tau,$$

in particular for (20):

$$\eta^*(x) = \int_0^1 F(x; \tau)(1 - \tau)\, d\tau.$$

η is a scalar potential and

$$h = \text{rot } j = [\text{grad } \eta^*, x]$$

is a vector potential of f:

$$\text{grad } \eta = f, \qquad \text{rot } h = f.$$

Notice that our vector potential has the additional property

$$\text{div } h = 0$$

and that its radial component vanishes. The equations

$$\text{rot } h = \text{grad } \eta, \qquad \text{div } h = 0,$$

which the reader is asked to compare with (11), show that $\Delta h = 0$: both scalar and vector potentials are harmonic.

We repeat the content of the first two equations of Theorem II. Any irrotational field f of finite $\|f\|$ splits uniquely into two orthogonal components,

$$f = g + e, \tag{21}$$

where e lies in $\mathfrak{E}$ and can therefore be expressed locally by a harmonic scalar potential η,

$$e = \text{grad } \eta,$$

while the $\mathfrak{G}$-component g is the limit (in the sense of the metric $\|f\|^2$) of a sequence of fields of the form

$$g_\nu = \text{grad } \psi_\nu \qquad (\psi_\nu \text{ of class } \Gamma;\ \nu = 1, 2, \cdots). \tag{22}$$

A similar decomposition (21) takes place for any solenoidal field f. We then express the $\mathfrak{E}$-component e by its harmonic vector potential h,

$$e = \text{rot } h \qquad (\text{div } h = 0),$$

while the $\mathfrak{G}'$-component g is the limit of a sequence of fields

$$g_\nu = \text{rot } v_\nu \qquad (v_\nu \text{ of class } \Gamma;\ \nu = 1, 2, \cdots). \tag{23}$$

In both cases we have

$$\|f\|^2 = \|g\|^2 + \|e\|^2.$$

Any properties of regularity prevailing for f (e.g., differentiability up to a certain order) will be shared by g because of the completely regular behavior of the component e.

We now are in a position to formulate the *problem of periods*. The integral of $(e \cdot dx)$ extended over a one-dimensional cycle (polygon) C in G is called the *period* $\pi(C)$. This period is a topological invariant of e, inasmuch as its value does not change under continuous deformation of C; this is a consequence of the equation div $e = 0$. In the solution (21) of our first boundary problem the component g is the limit of fields (22) whose 1-periods vanish. It is therefore reasonable to ascribe to f the same periods as to e. However, if f is continuous, we can form the integral of $(f \cdot dx)$ over C, and the question arises whether these natural periods of f coincide with those of e. Now let C^2 be a two-dimensional (oriented) cycle in G consisting of plane triangular faces. Because of rot $e = 0$ the surface integral $\int e_n\, do$ of the normal component of e over C^2 is a topological invariant $\pi(C^2)$, called two-dimensional period. In our second boundary problem the component g is the limit of fields (23) whose 2-periods vanish. Hence the 2-periods of e may be justly assigned to the solenoidal field f. Again, if f itself is continuous, one must ask whether its natural 2-period,

$$\pi(C^2) = \int f_n\, do \quad \text{over } C^2,$$

coincides with that of e.

3. A survey of vector analysis.[4] In order to prepare the answer to these questions we shall in this section concern ourselves with *continuous* vector fields f. The words "cube", "block", "square", "rectangle" will be used to designate cubes, parallelepipeds, squares and rectangles whose edges are parallel to the axes of coördinates. We add the adjective "oblique" to indicate arbitrary orientation of these figures in space.

The continuous vector field f in G is said to be *whirl-free* if to every point x^0 in G one can assign a neighborhood N (cube centered in x^0) such that the line integral

$$\int (f \cdot dx) = \int f_s \, ds = 0 \tag{24}$$

when extended over the boundary Q' of any square Q in N. Here s denotes the length along Q' and f_s the component of f in the direction of the line of integration. The equation (24) then holds for any square Q in G, irrespective of its size, as follows readily by subdividing Q into small squares. The same is true even for any rectangle R. Let us describe its edges of lengths a and b as horizontal and vertical. Divide the horizontal edge into a large number m of equal parts and fill the rectangle with squares of edge $\epsilon = a/m$ starting from the bottom. The line integral around each of these little squares vanishes. A small strip of m rectangles $r_1, \cdots, r_m$ of height $< \epsilon$ along the upper edge of R will be left over. Consider one of them, e.g. r_1, choose a point in r_1 and denote by f^1 the value of f at that point. If δ is an arbitrarily small given positive number, one can choose m so large that the inequality $|f - f^1| < \delta$ holds throughout r_1; on account of the *uniform* continuity of f this is even possible simultaneously for $r_1, \cdots, r_m$. Because the circumference of r_1 has a length $<4\epsilon$, one finds that the line integral of $f - f^1$ around r_1 has an absolute value $<4\epsilon \cdot \delta$. The line integral of the constant vector f^1 vanishes. Hence our estimate holds good for f itself, and the sum of the line integrals of f around $r_1, \cdots, r_m$ has an absolute value $<\delta \cdot 4\epsilon m = 4a\delta$. By adding all the little pieces into which R is cut, we find the same estimate holding for the line integral of f around R, and thus our statement is proved. This typical argument is very useful for a rigorous foundation of vector analysis.

Consider a cube T in G whose center is taken as the origin. We construct the solution φ in T of the equations

$$\frac{\partial \varphi}{\partial x_i} = f_i \qquad (i = 1, 2, 3) \tag{25}$$

[4] Here I follow closely the lectures on vector analysis which I used to give at the Technische Hochschule in Zürich; reflections of the method can be seen in some of my early papers, as in *Über die Randwertaufgabe der Strahlungstheorie und asymptotische Spektralgesetze*, Journal für Mathematik, vol. 143(1913), p. 182, footnote, and Sitzungsber. Preuss. Akad. d. Wissensch., 1917, p. 265.

which assumes a given value φ^0 at the origin by integrating parallel to the axes of coördinates. In the present general case this is the more natural procedure while we used radial integration for the harmonic fields in §2. We find

$$\varphi(x_1, x_2, x_3) = \varphi^0 + \int_0^{x_1} f_1(\xi, 0, 0)\, d\xi + \int_0^{x_2} f_2(x_1, \xi, 0)\, d\xi + \int_0^{x_3} f_3(x_1, x_2, \xi)\, d\xi.$$

By definition

$$\varphi(x_1, x_2, x_3 + \delta x_3) - \varphi(x_1, x_2, x_3) = \int_{x_3}^{x_3+\delta x_3} f_3(x_1, x_2, \xi)\, d\xi.$$

The analogous equations for the variation of x_2 and x_1 hold by definition for $x_3 = 0$ and for $x_3 = 0$, $x_2 = 0$ respectively, but using the fact that the integral of f around the circumference of a rectangle vanishes, we obtain them without those restrictions and thus prove (25). Once (25) is established one realizes that the line integral of f along any closed rectifiable curve in T, in particular along the circumference of a plane triangle or an oblique square, vanishes. The last remark shows that our definition of a whirl-free field is independent of the orientation of our Cartesian coördinate system. Summing over a finite set of triangles, each of which is embedded in a cube $T \subset G$, we see that the line integral of f over any 1-cycle C in G vanishes provided $C \sim 0$. As mentioned before, we assume our 1-chains, 1-cycles, 2-chains and 2-cycles to consist of straight segments and triangular faces respectively. $C \sim 0$ indicates that C bounds a certain 2-chain in G. By subdivision one can always take care that each of the faces of a given 2-chain can be immersed in a cube $T \subset G$.

It is convenient to use the "universal Abelian covering manifold" $\bar{G}$ over G on which a cycle C of G is closed if and only if $C \sim 0$ on G. The covering transformations S of $\bar{G}$ over G form an Abelian group. The process of continuation yields a solution φ of the equations (25),

$$\text{grad}\, \varphi = f,$$

over the whole of $\bar{G}$. The potential φ is periodic, the difference $\varphi(Sx) - \varphi(x)$ of the values of φ at any two points x, Sx of $\bar{G}$, covering the same point of G and arising from each other by the covering transformation S, is a constant π_S. If by following the 1-cycle C a traveler is led from a point x^0 to the point Sx^0 over x^0, this period π_S is the line integral of f over C.

The continuous field f in G is said to be *source-free* if one can assign a neighborhood $N \subset G$ to every point x^0 in G such that the flow

$$\mathfrak{F}[T] = \int f_n\, do$$

vanishes through the boundary T' of any cube T in N. By subdivision into sufficiently small cubes we carry this statement over to any cube $T \subset G$. How general is this law of vanishing flow? Consider a piece V of G bounded by one

or several surfaces V'. We cast over the space a net of cubes of small edge ϵ and thus cut V into small pieces. The flow for each cube of the net which lies entirely in V (inner piece) vanishes. To each of the truncated boundary pieces we apply the same method outlined before in the case of the small boundary rectangles r_i. By summation we find that the flow

$$\mathfrak{F}[V] = \int_{V'} f_n \, do$$

vanishes provided

(i) we may be sure that the flow of a *constant* vector f^0 vanishes for each boundary piece (projection in the direction of f^0!), and

(ii) the number of boundary pieces is $O(1/\epsilon^2)$.

This is certainly true for any solid convex polyhedron V which (together with its boundary) lies in G, in particular for a tetrahedron and any oblique cube. The last remark frees our basic definition from being bound to a special Cartesian coördinate system. Summing over a set of solid tetrahedrons, we realize that the flow through a 2-cycle C^2 vanishes provided $C^2 \sim 0$.

A device similar to what the universal Abelian covering manifold does for the whirl-free fields and their one-dimensional periods is missing in the present case.

Our next concern is the introduction of div and rot without the hypothesis of differentiability. We shall use the notations div*, rot* for these generalizations of the differential operators div, rot. The existence of

$$\operatorname{div}^* f = \rho$$

for a continuous field f means the existence of a continuous function ρ in G such that

$$\int_{T'} f_n \, do = \int_T \rho \qquad (T' \text{ boundary of } T) \tag{26}$$

holds for any cube T in a certain neighborhood N of any point x^0 of G. One sees at once that ρ is uniquely determined and that the restriction of T to a neighborhood N may be omitted. Moreover, the fundamental relation (26) may be carried over from T to any piece V of G which satisfies the two requirements (i), (ii) mentioned above; hence, in particular, to any tetrahedron and oblique block. Source-free fields f are characterized by the condition

$$\operatorname{div}^* f = 0.$$

The existence of

$$\operatorname{rot}^* f = u$$

for a continuous field f means the existence of a continuous vector field u in G such that

$$\int_Q u_n \, do = \int_{Q'} (f \cdot dx) \tag{27}$$

for any square Q in a certain neighborhood N of any point x^0 in G. The sense in which one travels around Q at the right side is to be connected with the normal n at the left by means of the orientation of space. u is uniquely determined. Whirl-free fields f are characterized by the equation

$$\mathrm{rot}^* f = 0.$$

In the manner described before the relation (27) carries over to any convex polygon parallel to one of the coördinate planes. However, we must try to make ourselves independent of this orientation. Consider any plane E. As it cannot be parallel to each of the three coördinate axes, one may suppose that it does not contain the "vertical" x_3-direction. Those parallelograms in E which by orthogonal projection upon the (x_1, x_2)-plane go into squares (in our limited sense) will for a moment be called cells of E. Suppose x^0 a point of G on E and $N \subset G$ a cube around the center x^0. Given any cell Z on E which lies in N, we project Z vertically upon the horizontal base of N and thus obtain a little chimney with a horizontal square base, four perpendicular walls which are parallel to the (x_1, x_3)- and the (x_2, x_3)-planes and the slanting opening Z. u satisfies the condition of vanishing flow

$$\int u_n\, do = 0$$

for the surface of each cube T and hence also for the surface of our chimney:

$$\sum_F \int_F u_n\, do + \int_Z u_n\, do = 0. \tag{28}$$

Since each of the five faces F (namely, the base and the four walls) have the proper orientation, we have for each of them the equation

$$\int_F u_n\, do = \int_{F'} f_s\, ds.$$

In view of (28) we obtain by summation over the five faces F:

$$-\int_Z u_n\, do = -\int_{Z'} f_s\, ds.$$

Operating with the cells Z in E as we operated with the squares in the planes parallel to the coördinate axes, we arrive at the equation

$$\int_P (\mathrm{rot}^* f)_n\, do = \int_{P'} f_s\, ds \tag{29}$$

for any plane convex polygon P in G. (This method is also well suited to carry Stokes' formula (29) over to curved surfaces.)

If f has continuous first derivatives, then $\mathrm{div}^* f$ and $\mathrm{rot}^* f$ arise from f by the differential operators div and rot respectively.

Just as, locally speaking, a whirl-free field f has a scalar potential φ, $f = \text{grad}\, \varphi$, so a source-free field f has a vector potential u,

$$f = \text{rot}^* u. \tag{30}$$

In general u will not be differentiable, and hence it is essential to write rot* instead of rot. We imitate the procedure followed for the scalar potential. The center of the cube T in G is taken as the origin. We choose $u_1 = 0$ along the x_1-axis, $u_2 = 0$ in the plane $x_3 = 0$, and $u_3 = 0$ throughout T. We integrate the equation

$$\frac{\partial u_2}{\partial x_1} - \frac{\partial u_1}{\partial x_2} = f_3$$

in the plane $x_3 = 0$ by

$$u_1(x_1, x_2, 0) = -\int_0^{x_2} f_3(x_1, \xi, 0)\, d\xi, \tag{31_0}$$

and

$$\frac{\partial u_3}{\partial x_2} - \frac{\partial u_2}{\partial x_3} = f_1, \qquad \frac{\partial u_1}{\partial x_3} - \frac{\partial u_3}{\partial x_1} = f_2$$

by

$$\begin{aligned} u_2(x_1, x_2, x_3) &= -\int_0^{x_3} f_1(x_1, x_2, \xi)\, d\xi, \\ u_1(x_1, x_2, x_3) &= u_1(x_1, x_2, 0) + \int_0^{x_3} f_2(x_1, x_2, \xi)\, d\xi. \end{aligned} \tag{31}$$

The relation

$$\int_Q f_n\, do = \int_{Q'} u_s\, ds \tag{32}$$

remains to be proved only for horizontal squares Q in T. Projecting such a Q perpendicularly upon the plane $x_3 = 0$, we obtain a block B for whose square base and four walls F the equation

$$\int_F f_n\, do = \int_{F'} u_s\, ds \tag{33}$$

holds by construction. Because f is supposed to be source-free, the total flow of f through the surface of the block, namely, through the five faces F and the square Q, vanishes. Hence summation of (33) over the five faces F yields the desired equation (32).

Our next step will be to carry over the formulas (13) and (2) to the starred operators. Here are the statements: If the function φ has continuous deriva-

tives and div* f exists for the continuous vector field f, then div* (φf) exists and is given by the equation

$$\text{div}^* (\varphi f) = \varphi \cdot \text{div}^* f + (f \cdot \text{grad } \varphi). \tag{34}$$

If f is continuous and rot* f exists while g is supposed to have continuous first derivatives, then

$$\text{div}^* [f, g] = (g \cdot \text{rot}^* f) - (f \cdot \text{rot } g). \tag{35}$$

Notice the asymmetry in the assumptions about f and g. In order to prove (34) cut the cube T into m^3 small cubes t of edge ϵ. Denote by φ^0 the value of any function φ in the center x^0 of t and by (ϵ) a quantity tending to zero with ϵ uniformly for all the m^3 cubes t. We have in t:

$$\varphi = \varphi^0 + \varphi_1^0(x_1 - x_1^0) + \varphi_2^0(x_2 - x_2^0) + \varphi_3^0(x_3 - x_3^0) + (\epsilon) \cdot \epsilon,$$

where $\varphi_i = \partial\varphi/\partial x_i$; hence, with an error of order $(\epsilon) \cdot \epsilon^3$,

$$\int_{t'} \varphi f_n \, do = \varphi^0 \cdot \int_{t'} f_n \, do + \int_{t'} \sum_i \varphi_i^0(x_i - x_i^0) \cdot f_n \, do.$$

In the first part we write

$$\int_{t'} f_n \, do = \int_t \text{div}^* f = \{(\text{div}^* f)^0 + (\epsilon)\} \cdot \epsilon^3.$$

In the second part we may replace f by the constant vector f^0 in neglecting something of the order $(\epsilon) \cdot \epsilon^3$, and then we readily obtain as its value

$$\epsilon^3 \cdot \sum_i \varphi_i^0 f_i^0 = \epsilon^3 \cdot (f \cdot \text{grad } \varphi)^0.$$

Hence

$$\int_{t'} \varphi f_n \, do = \epsilon^3 \{(\varphi \cdot \text{div}^* f + (f \cdot \text{grad } \varphi))^0 + (\epsilon)\},$$

and after summation over all m^3 cubes t:

$$\int_{T'} \varphi f_n \, do = \sum (\varphi \cdot \text{div}^* f + (f \cdot \text{grad } \varphi))^0 \cdot \epsilon^3 + (\epsilon).$$

In the limit $\epsilon \to 0$ the right member changes into

$$\int_T (\varphi \cdot \text{div}^* f + (f \cdot \text{grad } \varphi)).$$

Similarly (35) is proved.

If f is source-free and ψ of class Γ the equation (34) reduces to

$$\text{div}^* (\psi f) = (f \cdot \text{grad } \psi), \tag{36}$$

and thus by integration over G:

$$(4) \qquad \int (f \cdot \operatorname{grad} \psi) = 0.$$

Indeed, let G^* be the compact subset of G outside of which ψ vanishes. We construct a net of cubes of such width that all cubes of the net which have points in common with G^* lie in G. Integrate (36) over each of these cubes t,

$$\int_t (f \cdot \operatorname{grad} \psi) = \int_{t'} \psi f_n \, do$$

and form the sum. In the same way we infer from (35) that a whirl-free field satisfies the equation

$$\operatorname{div}^* [v, f] = (f \cdot \operatorname{rot} v)$$

and hence

$$(3) \qquad \int (f \cdot \operatorname{rot} v) = 0$$

for any vector field v of class Γ. In other words, *any source-free field f is solenoidal*, equation (4), *and any whirl-free field f is irrotational*, equation (3). *Lemma* 1 *is a particular case of both these facts.*

We cannot be fully satisfied without convincing ourselves of the inverse proposition: *A continuous field which is irrotational or solenoidal is whirl-free or source-free respectively.* I indicate the proof briefly in the two-dimensional case. Let Q be a square inside the "horizontal" two-dimensional region G. We use a third vertical dimension z and erect over Q a straight pyramid which we cut at the height $z = \epsilon$. Let the frustum, a low mound standing over Q, be described by the equation $z = \psi_\epsilon(x_1, x_2)$ while $\psi_\epsilon = 0$ outside Q. The derivatives of ψ_ϵ have simple discontinuities. Nevertheless we are allowed to substitute ψ_ϵ for the function ψ in the equation (4) which characterizes the continuous vector field f as solenoidal. The limit of

$$\frac{1}{\epsilon} \cdot \int_Q (f \cdot \operatorname{grad} \psi_\epsilon)$$

for $\epsilon \to 0$ is the flow $\int f_n \cdot do$ through Q'.

Finally we prove the modified equation (7):

$$(37) \qquad \operatorname{grad} (\operatorname{div} f) - \Delta^* f = \operatorname{rot}^* (\operatorname{rot} f).$$

Under the assumption that f and $\operatorname{div} f$ have continuous first derivatives, we maintain that the existence of $\operatorname{rot}^* (\operatorname{rot} f)$ implies the existence of

$$\Delta^* f_i = \operatorname{div}^* (\operatorname{grad} f_i) \qquad [i = 1, 2, 3]$$

and vice versa. We consider the vertical component and erect over a horizontal square $Q \subset G$ in the plane $x_3 = x_3^0$ a vertical block $B \subset G$ of height h. Its two

horizontal faces will be called Q and Q_h. Integrate the vertical component of the left member of (37) over B. The first term contributes

$$(38)\qquad \int_B \frac{\partial}{\partial x_3}(\operatorname{div} f) = \left[\int \operatorname{div} f\cdot do\right]_Q^{Q_h} = \int_{Q_h} \operatorname{div} f\cdot do - \int_Q \operatorname{div} f\cdot do.$$

Observe that

$$\int_Q \operatorname{div} f\cdot do = \int_Q \frac{\partial f_3}{\partial x_3}\, do + \int_Q \left(\frac{\partial f_1}{\partial x_1} + \frac{\partial f_2}{\partial x_2}\right) do$$
$$= \int_Q \frac{\partial f_3}{\partial x_3}\, do + \int_{Q'} f_n\cdot ds.$$

The second term at the right is a line integral around the boundary Q' of Q. Reversing for this part the transformation (38) we see that the contribution under consideration equals

$$\int \frac{\partial f_3}{\partial n}\, do + \int\int \frac{\partial f_n}{\partial x_3}\cdot ds\cdot dx_3,$$

the first integral extending over the two bases Q and Q_h, the second over the four walls. The second term of the left member of (37) contributes

$$\int_B \Delta^* f_3 = \int_{B'} \frac{\partial f_3}{\partial n}\, do.$$

Subtraction yields the integral of

$$\frac{\partial f_n}{\partial x_3} - \frac{\partial f_3}{\partial x_n} = (\operatorname{rot} f)_s$$

over the four walls or the integral over z from x_3^0 to $x_3^0 + h$ of the line integral

$$\int (\operatorname{rot} f)_s\cdot ds$$

around the cross section $Q_z : x_3 = z$ of B. Divide by h and pass to the limit $h \to 0$.

Let ρ be a continuous function in G and $T \subset G$ a cube. We form the potential $\Phi(x)$ of the mass distribution with density $\rho(x)$ in T:

$$\Phi(x) = \int \frac{\rho(x')}{|x - x'|};$$

the integration with respect to x' runs over T. Inside T not only Φ is continuous, but also its derivatives:

$$F = \operatorname{grad} \Phi = -\int \frac{x - x'}{|x - x'|^3}\cdot \rho(x').$$

(Incidentally F satisfies a Hölder condition with arbitrary exponent < 1.) For any cube t in the interior of T one has

$$(39) \qquad \int_{t'} F_n \, do = -4\pi \int_t \rho.$$

Indeed the integral

$$\int_{t'} \frac{(x - x')_n}{|x - x'|^3} \, do$$

running with respect to x over the surface t' is the solid angle under which t' appears from x' and thus equals 4π or 0 according to whether the point x' lies inside or outside t'. This gives for the left side of (39) the result

$$-4\pi \cdot \int \rho(x') \qquad (x' \text{ over } t).$$

The formula (39) may be stated thus:

$$\operatorname{div}^* F = \Delta^* \Phi = -4\pi\rho \qquad \text{inside } T.$$

We return to an arbitrary source-free field f in G. The local existence of a vector potential u of f, equation (30), is of little use unless we are able to normalize it by the equation $\operatorname{div}^* u = 0$. Here is the way in which this may be accomplished. We form for any point x in T the integral

$$j(x) = \frac{1}{4\pi} \int \frac{f(x')}{|x - x'|} \qquad (x' \text{ over } T).$$

Then

$$(40) \qquad \Delta^* j = -f \qquad \text{in } T.$$

On the other hand,

$$\operatorname{div} j = -\frac{1}{4\pi} \cdot \int \frac{(x - x') \cdot f(x')}{|x - x'|^3} \qquad (x' \text{ over } T).$$

Write for a moment x^0, x instead of x, x'; the value $(\operatorname{div} j)^0$ at x^0 is given by

$$-\frac{1}{4\pi} \int_T \left(\operatorname{grad} \frac{1}{|x^0 - x|} \cdot f(x) \right).$$

Use the relation (34) for

$$\varphi(x) = \frac{1}{|x^0 - x|}$$

and integrate over T after first excluding x^0 by a small cubical neighborhood t. Taking account of $\operatorname{div}^* f = 0$ and shrinking t down to x^0, we get

$$(\operatorname{div} j)^0 = -\frac{1}{4\pi} \int_{T'} \frac{1}{|x^0 - x|} f_n(x)$$

or

$$\theta = -\operatorname{div} j = \frac{1}{4\pi} \int \frac{f_n(x')}{|x - x'|} \qquad (x' \text{ over } T').$$

This θ is a harmonic function in T with derivatives of all orders, and thus we deduce from (40) and (37) the relation

$$\operatorname{rot}^* (\operatorname{rot} j) = f - \operatorname{grad} \theta.$$

We saw at the end of §2 how we may ascertain a harmonic vector field h in T such that

$$\operatorname{grad} \theta = \operatorname{rot} h, \qquad \operatorname{div} h = 0.$$

The field

$$\operatorname{rot} j + h = u$$

satisfies all our demands. This construction is superior to the one laid down in the equations (31), and the result, which is the main goal of this long section, deserves to be fixed as

THEOREM IV. *For a harmonic function η in a cube T we can find a harmonic vector field h such that*

$$\operatorname{grad} \eta = \operatorname{rot} h, \qquad \operatorname{div} h = 0.$$

For any source-free vector field f in T we can find a continuous field u in T (which moreover satisfies a Hölder condition of arbitrary exponent < 1) *such that*

$$f = \operatorname{rot}^* u, \qquad \operatorname{div}^* u = 0. \tag{41}$$

4. **Auxiliary inequalities.** We now come to another more interesting preparation. Take a cube T of edge l and let it be described by the inequalities

$$0 \leqq x_1 \leqq l, \qquad 0 \leqq x_2 \leqq l, \qquad 0 \leqq x_3 \leqq l.$$

Let φ be a function with continuous first derivatives in T (boundary included). The gradient $f = \operatorname{grad} \varphi$ determines φ up to an additive constant. We therefore split φ into a multiple φ^0 of 1 and a second component φ^* orthogonal to 1,

$$\varphi(x) = \varphi^0 + \varphi^*(x), \qquad \int_T \varphi^*(x) = 0$$

and set

$$H_T[\varphi] = \int_T (\varphi^*)^2, \qquad D_T[\varphi] = \int_T (\operatorname{grad} \varphi)^2.$$

LEMMA 3. *With a certain constant L depending only on l, Poincaré's inequality*[5]

$$H_T[\varphi] \leqq L^2 \cdot D_T[\varphi] \tag{42}$$

[5] H. Poincaré, *Sur les équations de la physique mathématique*, Rend. Circ. Mat. Palermo, vol. 8(1894), pp. 70–76.

holds. The best value for L is

$$L = \frac{l}{\pi}.$$

The attempt to minimize $D_T[\varphi]$ under the auxiliary condition $\int_T \varphi^2 = 1$ leads to the acoustic problem of the eigen-tones of T, even when T is any fairly regular region:

$$\Delta\varphi + \lambda^2\varphi = 0 \quad \text{in } T, \qquad \frac{\partial\varphi}{\partial n} = 0 \quad \text{on the boundary } T'.$$

Hence the best value for L in (42) is the reciprocal of the frequency λ of the gravest eigen-tone. This observation points the way to the following proof.

Take $l = 1$. For any set $\nu = (\nu_1, \nu_2, \nu_3)$ of non-negative integers we introduce the eigen-function

$$\text{co}(\nu) = \text{co}(\nu_1, \nu_2, \nu_3) = \cos \pi\nu_1 x_1 \cdot \cos \pi\nu_2 x_2 \cdot \cos \pi\nu_3 x_3$$

and the eigen-vectors

$$\text{co}^{(1)}(\nu), \text{co}^{(2)}(\nu), \text{co}^{(3)}(\nu); \qquad \text{sn}^{(1)}(\nu), \text{sn}^{(2)}(\nu), \text{sn}^{(3)}(\nu).$$

The second and third component of $\text{co}^{(1)}(\nu)$ and $\text{sn}^{(1)}(\nu)$ vanish while their first components equal

$$\sin \pi\nu_1 x_1 \cdot \cos \pi\nu_2 x_2 \cdot \cos \pi\nu_3 x_3 \quad \text{and} \quad \cos \pi\nu_1 x_1 \cdot \sin \pi\nu_2 x_2 \cdot \sin \pi\nu_3 x_3$$

respectively. We introduce the Fourier coefficients of φ and $f = \text{grad } \varphi$ with respect to $\text{co}(\nu)$ and $\text{co}^{(1)}(\nu), \text{co}^{(2)}(\nu), \text{co}^{(3)}(\nu)$ respectively:

$$a = a(\nu) = \int_T \varphi \cdot \text{co}(\nu), \qquad a_i = a_i(\nu) = \int_T f \cdot \text{co}^{(i)}(\nu) \qquad (i = 1, 2, 3).$$

φ^* has the same Fourier coefficients as φ, except that $a(0, 0, 0)$ is replaced by 0. Set

$$2(\nu) = 1 \text{ or } 2 \text{ for } \nu = 0 \text{ or } \nu = 1, 2, \cdots \text{ respectively,}$$

and

$$\delta = \delta(\nu_1, \nu_2, \nu_3) = 2(\nu_1) \cdot 2(\nu_2) \cdot 2(\nu_3).$$

The individual eigen-function $\text{co}(\nu)$ contributes to $H_T[\varphi]$ the amount $\delta \cdot a^2$; according to Parseval's equation, $H_T[\varphi]$ itself is the sum of these contributions associated with the different sets $\nu = (\nu_1, \nu_2, \nu_3)$, excluding $(0, 0, 0)$.

The connection between φ^* and f may be put in this form:

$$\int_T \varphi^* = 0,$$

$$\int_T \varphi^* \cdot \text{div } p = -\int_T (f \cdot p),$$

the second equation holding for any vector field p in T with continuous derivatives whose normal component p_n vanishes along T'. In the sequel p always has this significance. Taking

$$p = \mathrm{co}^{(1)}, \mathrm{co}^{(2)}, \mathrm{co}^{(3)}$$

one finds

$$a_i = -\pi \nu_i a \qquad (i = 1, 2, 3).$$

Since

$$\pi^2 \cdot \sum_\nu{}' (\nu_1^2 + \nu_2^2 + \nu_3^2) \cdot \delta a^2 = \sum_\nu{}' \delta \cdot (a_1^2 + a_2^2 + a_3^2) \leqq \int_T f^2,$$

where the accent indicates the exclusion of (0, 0, 0) from summation, we obtain the desired result

$$\pi^2 \cdot H_T[\varphi] \leqq \int_T f^2.$$

As one sees, the proof depends on the fact that

$$[2(\nu)]^{\frac{1}{2}} \cdot \cos \pi \nu x \qquad (\nu = 0, 1, 2, \cdots)$$

constitute a complete orthogonal system for the interval $0 \leqq x \leqq 1$.

For a block T of arbitrary edges l_1, l_2, l_3,

$$0 \leqq x_1 \leqq l_1, \quad 0 \leqq x_2 \leqq l_2, \quad 0 \leqq x_3 \leqq l_3,$$

one would obtain the same inequality (42) with

$$L = \frac{1}{\pi} \max (l_1, l_2, l_3). \tag{43}$$

The inequality (42) may also be written in this form:

$$\int\int (\varphi(x) - \varphi(x'))^2 \leqq 2L^2 T \cdot D_T[\varphi]. \tag{44}$$

In the left member both integrations with respect to x and x' run over T. At the right, the letter T stands for the volume of the cube T.

What is the analogous relation for rot instead of grad? rot u remains unchanged by adding a term grad φ to u. Hence as a preliminary we carry out the following construction[6] in which T plays the rôle of G. Let $\mathfrak{T}$ be the complete Hilbert space of all vector fields u of finite

$$\| u \|_T^2 = \int_T u^2$$

[6] This is the problem to which O. Nikodym applied the method of orthogonal projection, loc. cit. (footnote 1).

while $\bar{\mathfrak{T}} \subset \mathfrak{T}$ is the closure of all fields grad σ derived from potentials σ that have continuous first derivatives in the closed T, and $\mathfrak{T}^*$ consists of the elements u^* of $\mathfrak{T}$ which are orthogonal to all those fields:

$$\int_T (u^* \cdot \operatorname{grad} \sigma) = 0. \tag{45}$$

This orthogonal decomposition $\mathfrak{T} = \mathfrak{T}^* + \bar{\mathfrak{T}}$ when applied to an element u of $\mathfrak{T}$ is described by

$$u = u^* + \bar{u}.$$

We put the relation $f = \operatorname{rot} u$ into the more universal form

$$\int_T (u \cdot \operatorname{rot} q) = \int_T (f \cdot q),$$

holding for any field q in the closed T which has continuous derivatives and whose tangential components vanish along T'. This equation survives for u^* because the fields grad σ approximating $\bar{u}$ and hence $\bar{u}$ itself are orthogonal to rot q. We thus arrive at this conjecture:

Lemma 4. *Let f and u^* be vector fields in the interior of T of finite $\|f\|_T$, $\|u^*\|_T$ such that the equations*

$$\int_T (u^* \cdot \operatorname{grad} \sigma) = 0, \qquad \int_T (u^* \cdot \operatorname{rot} q) = \int_T (f \cdot q) \tag{46}$$

hold whenever the scalar σ and vector q have continuous derivatives in the closed T and the tangential components of q vanish along the boundary of T. Then u^* is uniquely determined by f, and a universal inequality

$$\| u^* \|_T^2 \leqq M^2 \cdot \| f \|_T^2 \tag{47}$$

holds where the constant M depends on l only; its best value is

$$M^2 = \frac{l^2}{2\pi^2}.$$

The corresponding minimizing problem is that of radiation, of standing electromagnetic waves in the Hohlraum T. This sets the tune for our proof.

Let a_i, b_i $(i = 1, 2, 3)$ be the Fourier coefficients of f with respect to $\mathrm{sn}^{(i)}$ and of u with respect to $\mathrm{co}^{(i)}$. Then by the substitutions $\sigma = \mathrm{co}\,(\nu)$ and $q = \mathrm{co}^{(1)}, \mathrm{co}^{(2)}, \mathrm{co}^{(3)}$ the equations (46) yield

$$\begin{gathered} b_1\nu_1 + b_2\nu_2 + b_3\nu_3 = 0, \\ \pi(\nu_2 b_3 - \nu_3 b_2) = a_1, \quad \pi(\nu_3 b_1 - \nu_1 b_3) = a_2, \quad \pi(\nu_1 b_2 - \nu_2 b_1) = a_3, \end{gathered} \tag{48}$$

and from this follows

$$a_1^2 + a_2^2 + a_3^2 = \pi^2(\nu_1^2 + \nu_2^2 + \nu_3^2)(b_1^2 + b_2^2 + b_3^2).$$

Not only (0, 0, 0) but also those combinations (ν_1, ν_2, ν_3) in which two components vanish may be excluded. Indeed $\nu_2 = \nu_3 = 0, \nu_1 \neq 0$ implies $a_1 = a_2 = a_3 = 0$, $b_2 = b_3 = 0$, and by (48) also $b_1 = 0$. Thus the least admissible value of $\nu_1^2 + \nu_2^2 + \nu_3^2$ is 2. This proves our statement.

For a block T with the edges l_1, l_2, l_3 a similar inequality holds, and in that case the best value for M^2 is

$$M^2 = \frac{1}{\pi^2}\cdot\max\left(\frac{l_2^2 l_3^2}{l_2^2 + l_3^2}, \frac{l_3^2 l_1^2}{l_3^2 + l_1^2}, \frac{l_1^2 l_2^2}{l_1^2 + l_2^2}\right).$$

The question of the existence of u^* for a given solenoidal f will be decided in connection with the application of our inequality to the situation in which we are primarily interested (§6).

Similar inequalities hold for a sphere K of radius R. The best values of L and M in (42) and (47) are R/α, where α is the least positive root of the equation

$$\tan \alpha = \frac{2\alpha}{2 - \alpha},$$

and $2R/\pi$ respectively. Elementary proofs of the inequality (42) for convex bodies and sphere, which, however, do not yield the best constants, are given by H. Poincaré and the author.[7] It seems unlikely that they allow approaching the inequality (47).

5. Periods of irrotational fields.

1. We take up the first boundary problem, i.e., the decomposition $\mathfrak{F} = \mathfrak{G} + \mathfrak{E}$, $f = g + e$, of an element f of $\mathfrak{F}$. Here g is the limit of a sequence of fields $g_\nu = \operatorname{grad} \psi_\nu$. For the moment we throw aside our knowledge that ψ_ν is single valued over the whole of G and vanishes in a boundary strip. Given a cube $T \subset G$, we subtract from ψ_ν its mean value ψ_ν^0 over T, so that $\psi_\nu^* = \psi_\nu - \psi_\nu^0$ satisfies the normalizing condition

$$\int_T \psi_\nu^* = 0.$$

We apply Lemma 3:

$$\int_T (\psi_\nu^* - \psi_\mu^*)^2 \leqq L^2 \cdot \| g_\nu - g_\mu \|^2.$$

[7] H. Poincaré, loc. cit. (footnote 5). See also R. Courant and D. Hilbert, *Methoden der mathematischen Physik*, II, Berlin, 1937, pp. 488 and 517–519. H. Weyl, *Die Idee der Riemannschen Fläche*, Leipzig, 1913, pp. 89–90. Another proof could be modeled after the pattern of the argument in §7.

Hence the integral at the left tends to zero with $\mu, \nu \to \infty$, and there exists a function ψ^* of finite square integral in T such that

$$\int_T (\psi^* - \psi_\nu^*)^2 \to 0 \quad \text{for} \quad \nu \to \infty.$$

ψ^* is connected with g by the equations $\int_T \psi^* = 0$ and

$$\int_T \psi^* \cdot \operatorname{div} p = -\int_T (g \cdot p) \tag{49}$$

holding for any p of the formerly described nature. If one writes $e = \operatorname{grad} \eta$, then $\psi^* + \eta$ stands in similar relationship to f itself.

2. We study any function φ in G with continuous derivatives and the accompanying vector field $f = \operatorname{grad} \varphi$. We suppose

$$\|f\|^2 = \int_G f^2 = D[\varphi] \leqq \epsilon^2. \tag{50}$$

Choosing a definite cube T_0 of edge l_0 in G, we normalize the arbitrary additive constant in φ by

$$\int_{T_0} \varphi = 0, \tag{51}$$

and then have

$$\int_{T_0} \varphi^2 \leqq L_0^2 \epsilon^2 \qquad (L_0 = l_0/\pi). \tag{52}$$

Lemma 5. *Let T, T_1 be any two overlapping cubes in G. We maintain that under the assumption* (50)

$$\int_T \varphi^2 \leqq A^2 \epsilon^2 \quad \textit{implies} \quad \int_{T_1} \varphi^2 \leqq A_1^2 \epsilon^2, \tag{53}$$

where A_1 depends on A and T, T_1 but not on ϵ and φ.

Proof. According to Lemma 3 there exists a constant φ^1, the mean value of φ over T_1, such that

$$\int_{T_1} (\varphi - \varphi^1)^2 \leqq L_1^2 \epsilon^2 \qquad (L_1 = l_1/\pi). \tag{54}$$

Denote by t the common part of T and T_1. The hypothesis (53) and formula (54) imply

$$\int_t \varphi^2 \leqq A^2 \epsilon^2, \qquad \int_t (\varphi^1 - \varphi)^2 \leqq L_1^2 \epsilon^2.$$

Using Schwarz's inequality for

$$\varphi^1 = A^{\frac{1}{2}} \cdot \frac{\varphi}{A^{\frac{1}{2}}} + L_1^{\frac{1}{2}} \cdot \frac{\varphi^1 - \varphi}{L_1^{\frac{1}{2}}}$$

we find

$$(\varphi^1)^2 t \leqq (A + L_1)^2 \epsilon^2.$$

Since

$$\int_{T_1} \varphi^2 = \int_{T_1} (\varphi - \varphi^1)^2 + (\varphi^1)^2 T_1,$$

we arrive at the result (53) with

$$A_1^2 = L_1^2 + (A + L_1)^2 \cdot \frac{T_1}{t}.$$

Let T be any cube in G. Join T_0 with T by a chain of cubes any two consecutive members of which overlap.[8] (We operate here, if necessary, in one of the connected components of G, rather than in G itself.) Starting with (52) and continuing by means of the last lemma, we ascertain a constant A_T such that

$$\int_T \varphi^2 \leqq A_T^2 \cdot \epsilon^2 \tag{55}$$

follows from

$$\int_{T_0} \varphi = 0 \quad \text{and} \quad D[\varphi] \leqq \epsilon^2.$$

The application to our situation is immediate. We subtract from ψ_ν its mean value ψ_ν^0 over T_0, so that $\psi_\nu^* = \psi_\nu - \psi_\nu^0$ fulfills the normalizing condition

$$\int_{T_0} \psi_\nu^* = 0. \tag{56}$$

Then

$$\int_T (\psi_\nu^* - \psi_\mu^*)^2 \leqq A_T^2 \cdot || g_\nu - g_\mu ||^2$$

for any cube $T \subset G$. Hence we find a function ψ^* in the whole of G such that

$$\int_T (\psi^* - \psi_\nu^*)^2 \to 0 \quad \text{for each } T. \tag{57}$$

The relation (49) now holds for every T. The harmonic e has a harmonic scalar potential η, $e = \text{grad } \eta$, on $\bar{G}$. We put $\varphi^* = \psi^* + \eta$ and feel justified in

[8] This argument is taken from the author's book on Riemann surfaces, loc. cit. (footnote 7), pp. 103–104.

ascribing the value $\varphi^*(b) - \varphi^*(a)$ to the line integral $\int (f \cdot dx)$ extending over any line $\mathfrak{l}$ on G joining the point a with b. Here a denotes at the same time a point on $\bar{G}$ lying over a and b that point on $\bar{G}$ to which a traveler starting at a is led by following the trace $\mathfrak{l}$.

3. Suppose f itself to be continuous and thus whirl-free rather than irrotational. We can write $f = \text{grad}\ \varphi$, where φ is a periodic function on $\bar{G}$ with continuous derivatives. Our last result comes very close to a proof that φ has the same periods as η. But to make this point quite explicit we apply Lemma 5 to our present φ on $\bar{G}$, which we are evidently entitled to do, and to a chain leading from the cube T_0 on $\bar{G}$ to that cube T which arises from T_0 by the covering transformation S. We make use of the normalization (51). The inequalities

$$\int_{T_0} \varphi^2 \leqq L_0^2 \epsilon^2, \qquad \int_T \varphi^2 = \int_{T_0} (S\varphi)^2 \leqq A_T^2 \cdot \epsilon^2$$

result in the following relation for the constant difference $\pi_S = S\varphi - \varphi$:

$$\pi_S^2 \cdot T_0 \leqq (L_0 + A_T)^2 \epsilon^2.$$

Hence:

Theorem V. *For any covering transformation S of $\bar{G}$ there exists a constant H_S such that the period $\pi_S[f]$ of any whirl-free field f satisfies the inequality*

$$\pi_S^2 \leqq H_S^2 \cdot \|f\|^2.$$

We apply this lemma to $g - g_\nu$ with its potential $\varphi - \eta - \psi_\nu$. Its periods are independent of ν, namely, the periods κ_S of $\varphi - \eta$, and we obtain

$$\kappa_S^2 \leqq H_S^2 \cdot \|g - g_\nu\|^2,$$

and this proves the desired result $\kappa_S = 0$:

Theorem VI. *If the decomposition $\mathfrak{F} = \mathfrak{G} + \mathfrak{E}$ is applied to a whirl-free vector f, $f = \text{grad}\ \varphi$, then the periods of its potential φ agree with those of the harmonic potential η of the $\mathfrak{E}$-part e.*

The function $\psi^* = \varphi - \eta$, single-valued in G, is the limit of the functions ψ_ν^* in the sense of the relation (57), provided the arbitrary additive constant in η is normalized according to

$$\int_{T_0} \psi^* = 0.$$

Our theorem proves in particular that for any periodic function φ on $\bar{G}$ with continuous derivatives we can ascertain a harmonic function η with the same periods; it goes beyond this statement by adding that in some sense η has the same boundary values as φ.

6. **Periods of solenoidal fields.** We shall try to imitate as closely as possible the procedure of the previous section for the second boundary problem, i.e., the

decomposition

$$\mathfrak{F}' = \mathfrak{G}' + \mathfrak{E}, \quad f = g + e$$

of a solenoidal vector field f. We encounter no difficulty at point 1. g is the limit of fields $g_\nu = \text{rot } v_\nu$, and we apply to v_ν and a given cube $T \subset G$ the splitting formerly described as $\mathfrak{T} = \mathfrak{T}^* + \bar{\mathfrak{T}}$. Lemma 4 yields

$$\int_T (v_\nu^* - v_\mu^*)^2 \leqq M^2 \cdot \| g_\nu - g_\mu \|^2,$$

and hence v_ν^* tends to a limit v^* in T in the sense that

$$\int_T (v^* - v_\nu^*)^2 \to 0 \quad \text{for } \nu \to \infty.$$

This v^* satisfies the relations

$$\int_T (v^* \cdot \text{grad } \sigma) = 0,$$

$$\int_T (v^* \cdot \text{rot } q) = \int_T (g \cdot q)$$

with the same conditions for σ and q as in (46). Put e in the form rot h; then the given field f will stand in the same relationship to $v^* + h^*$ as g to v^*. (Incidentally h^* would equal h if we had used a sphere K instead of a cube T, because div h and the radial component of h vanish.)

We can expect no analogue of point 2.

Point 3 may be established although the argument has to undergo some modification on account of the missing parallel to the covering manifold $\bar{G}$.

Consider a 2-cycle C in G whose faces are plane triangles Δ each enclosed in a cubical box $T \subset G$. (To be precise, we suppose all points of the closed triangle to lie in the interior of T.) Our goal is to derive an inequality for the period

$$\pi_C = \int_C f_n \, do$$

of any source-free vector field f in G,

$$\pi_C^2 \leqq H_C^2 \cdot \| f \|^2$$

in which the constant H_C depends on C and not on f. To that end we enclose T in a slightly larger cube $T' \subset G$ and construct a vector potential u for f, continuous in the interior of T' and satisfying the relations

(58) $$f = \text{rot}^* u, \qquad \text{div}^* u = 0.$$

Afterwards we pass to T and apply the splitting $\mathfrak{T} = \mathfrak{T}^* + \bar{\mathfrak{T}}$ to u,

$$u = u^* + \bar{u}.$$

We maintain that $\bar{u} = \text{grad}\, \vartheta$, where ϑ is harmonic in the interior of T. Indeed, by definition,

$$\int_T (\bar{u} \cdot \text{rot}\, q) = 0.$$

Moreover, the second equation (58) implies

$$\int_T (u \cdot \text{grad}\, \psi) = 0$$

for any ψ with continuous derivatives in the closed T which vanishes at the border of T, and this, together with (45), results in

$$\int_T (\bar{u} \cdot \text{grad}\, \psi) = 0.$$

Our fundamental Theorem I applied to T instead of G then shows that $\bar{u}$ is a harmonic field satisfying the relations

$$\text{div}\, \bar{u} = 0, \quad \text{rot}\, \bar{u} = 0.$$

Hence u^* is continuous in the interior of T and the relations (58) persist for u^*. Therefore

$$\int_\Delta f_n\, do = \int_{\Delta'} u_s\, ds = \int_{\Delta'} u_s^*\, ds. \tag{59}$$

Since the conditions (46) prevail, we conclude from Lemma 4 that

$$\int_T (u^*)^2 \leqq M^2 \cdot \|f\|^2. \tag{60}$$

[The combinatorial schemes of the 2-cycles I had in mind when writing this proof are abstract finite oriented 2-*manifolds*. However, with obvious alterations the proof holds good for 2-cycles on an arbitrary 2-*complex*, and Theorems VII and VIII should be understood in this wider sense. **Added November 17, 1940.**]

Next we envisage two adjacent faces Δ_1, Δ_2 of C, their common edge δ, and their embedding boxes T_1, T_2 in which the fields u_1^*, u_2^* are defined. In computing π_C we sum (59) over all faces of C. The edge δ contributes

$$\int_\delta (u_1^* - u_2^*)_s\, ds. \tag{61}$$

T_1 and T_2 have a common part T_δ. The boxes T for all the faces Δ with a common vertex a have a common part T_a. The T_δ and T_a are not cubes, but blocks. Because

$$f = \text{rot}^*\, u_1^* = \text{rot}^*\, u_2^* \quad \text{in } T_\delta$$

the difference $u_2^* - u_1^*$ is whirl-free in T_δ and hence equals grad φ_δ, where $\varphi_\delta = \varphi_{12}$ is a function in the interior of T_δ with continuous derivatives. Incidentally φ_δ is harmonic because $u_2^* - u_1^*$ is not only whirl-free but also source-free according to the equations

$$\operatorname{div}^* u_1^* = 0, \qquad \operatorname{div}^* u_2^* = 0.$$

The relations (60) for u_1^* in T_1 and u_2^* in T_2 imply

$$\int_{T_\delta} (u_1^*)^2 \leqq M_1^2 \cdot \|f\|^2, \qquad \int_{T_\delta} (u_2^*)^2 \leqq M_2^2 \cdot \|f\|^2$$

and hence

$$\int_{T_\delta} (\operatorname{grad} \varphi_\delta)^2 = \int_{T_\delta} (u_1^* - u_2^*)^2 \leqq (M_1 + M_2)^2 \cdot \|f\|^2. \tag{62}$$

Let the edge δ lead from the vertex a to b. Its orientation is such that the transition $\Delta_1 \to \Delta_2$ crosses it from left to right. For the opposite crossing $\Delta_2 \to \Delta_1$ we obtain $\varphi_{21} = -\varphi_{12}$, $\varphi_{-\delta} = -\varphi_\delta$. The integral (61) has the value

$$-\int_\delta (\operatorname{grad} \varphi_\delta)_s \, ds = \varphi_\delta(a) - \varphi_\delta(b) = \varphi_\delta(a) + \varphi_{-\delta}(b).$$

Summing over all edges δ of C we arrive at this formula:

$$\pi_C = \sum_\delta \{\varphi_\delta(a) - \varphi_\delta(b)\} = \sum_a \chi_a(a). \tag{63}$$

The latter sum extends over all vertices a and

$$\chi_a(a) = \sum_\delta{}' \varphi_\delta(a),$$

where $\sum_\delta{}'$ runs over all edges δ radiating from a.

Denote by $\Delta_1, \Delta_2, \cdots, \Delta_m$ (Δ_1) the cycle of faces around a. Because

$$\operatorname{grad} \varphi_{12} = u_2^* - u_1^*, \quad \operatorname{grad} \varphi_{23} = u_3^* - u_2^*, \quad \cdots$$

the gradient of the sum $\varphi_{12} + \varphi_{23} + \cdots + \varphi_{m1}$ vanishes in T_a, and hence that sum is a constant. Designating by x_a a point varying over T_a we thus find that

$$\chi_a(x_a) = \sum_\delta{}' \varphi_\delta(x_a)$$

is a constant, and (63) changes into the more general equation

$$\pi_C = \sum_a \chi_a(x_a) = \sum_\delta \{\varphi_\delta(x_a) - \varphi_\delta(x_b)\}. \tag{64}$$

This equation puts in evidence that the value of π_C is not affected by moving each vertex a within its surrounding box T_a.

We apply the inequality (44) to φ_δ and the box T_δ with the effect that

$$\iint (\varphi_\delta(x) - \varphi_\delta(x'))^2 \leqq L_\delta^2 \cdot \int_{T_\delta} (\operatorname{grad} \varphi_\delta)^2$$

with a certain constant L_δ depending on the lengths of the edges of T_δ only; see equation (43). Integration at the left ranges over T_δ with respect to x and x'. We combine this with (62) and limit the integration at the left for x to T_a and for x' to T_b and thus find an inequality

$$\int_{T_a}\int_{T_b} (\varphi_\delta(x_a) - \varphi_\delta(x_b))^2 \leqq A_\delta^2 T_a T_b \cdot ||f||^2.$$

The constant A_δ, whose explicit value can easily be given, does not depend on f. If we now use the second expression (64) of π_C and integrate by running each point x_a over its box T_a we get

$$\pi_C^2 \leqq H_C^2 \cdot ||f||^2 \quad \text{where } H_C = \sum_\delta A_\delta :$$

THEOREM VII. *To every 2-cycle C in G one can assign a constant H_C such that the period π_C over C of any source-free vector field f in G satisfies the inequality*

$$\pi_C^2 \leqq H_C^2 \cdot ||f||^2.$$

Returning to our second boundary problem we observe that the period κ_C over any 2-cycle C of g is that of

$$g - g_\nu = g - \text{rot } v_\nu .$$

The resulting inequality

$$\kappa_C^2 \leqq H_C^2 \cdot ||g - g_\nu||^2 \quad \text{for } \nu = 1, 2, \cdots$$

shows that $\kappa_C = 0$ or that f has the same periods as the harmonic part e:

THEOREM VIII. *If the decomposition $\mathfrak{F}' = \mathfrak{G}' + \mathfrak{E}$ is applied to a source-free vector field f, then the periods of f over any 2-cycle are the same as those of the $\mathfrak{E}$-part e.*

7. **Behavior on the boundary.** So far we have evaluated the fact that the approximating g_ν in our two problems have a scalar and a vector potential respectively, which are single-valued throughout G. How can we take into account their vanishing in a boundary strip?

In our first problem,

$$g_\nu = \text{grad } \psi_\nu \qquad (\psi_\nu \text{ of class } \Gamma),$$

we put $\psi_\nu = 0$ outside G, so that ψ_ν is a function with continuous derivatives throughout the whole space, and choose an arbitrarily large cube T which need not be part of G. We are going to prove that the ψ_ν themselves, without being normalized by subtraction of a constant to fit the equation (56), converge to a limit $\bar{\psi}$ in the sense that

$$\int_T (\bar{\psi} - \psi_\nu)^2 \to 0 \quad \text{for } \nu \to \infty.$$

To this end, study any function ψ of class Γ and the potential Φ of the mass distribution of density

$$\rho(x) = \psi(x) \text{ in } T, \qquad = 0 \text{ outside } T:$$

$$\Phi(x) = \int_{x'} \frac{\rho(x')}{|x - x'|},$$

and form the integral

$$J = \frac{1}{4\pi} \int (\text{grad } \Phi \cdot \text{grad } \psi).$$

Taking G as the domain of integration one finds

$$J = -\frac{1}{4\pi} \int \psi \cdot \Delta^* \Phi = \int \psi\rho = \int_T \psi^2. \tag{65}$$

Taking the whole space as domain of integration one gets

$$J^2 \leqq \frac{1}{4\pi} \int (\text{grad } \Phi)^2 \cdot \frac{1}{4\pi} \int (\text{grad } \psi)^2.$$

The first integral at the right converges because Φ, grad Φ vanish at infinity with the orders 1 and 2 respectively, and it equals

$$-\frac{1}{4\pi} \int \Phi \cdot \Delta^* \Phi = \int\int \frac{\rho(x)\rho(x')}{|x - x'|}.$$

At the right side the integration with respect to both x and x' may be restricted to T and then $\rho(x)$ be replaced by $\psi(x)$. By Schwarz's inequality its square is less than or equal to

$$E^2 \cdot \int_T \int_T (\psi(x)\psi(x'))^2 = E^2 \cdot \left(\int_T \psi^2\right)^2,$$

where

$$E^2 = \int_T \int_T \frac{1}{|x - x'|^2}.$$

Thus

$$J^2 \leqq \frac{E}{4\pi} \cdot \int_T \psi^2 \cdot \int_G (\text{grad } \psi)^2.$$

Combine this with (65). The result is the following inequality which ought to be compared with Lemma 3:

$$\int_T \psi^2 \leqq \frac{E}{4\pi} \cdot \int_G (\text{grad } \psi)^2.$$

Application to $\psi_\nu - \psi_\mu$ yields

$$\int_T (\psi_\nu - \psi_\mu)^2 \to 0 \quad \text{for } \mu, \nu \to \infty.$$

Hence ψ_ν converges toward a limit ψ in the sense that for every cube T

$$\int_T (\psi - \psi_\nu)^2 \to 0 \quad \text{for } \nu \to \infty.$$

ψ vanishes outside G. The relation

$$\int_T \psi_\nu \cdot \operatorname{div} p = -\int_T (g_\nu \cdot p)$$

holding for any p satisfying the conditions indicated by that letter carries over to the limit:

$$\int_T \psi \cdot \operatorname{div} p = -\int_T (g \cdot p). \tag{66}$$

Again g is set equal to zero outside G. This is the sought-for equivalent for the equation $g = \operatorname{grad} \psi$ together with the boundary condition $\psi = 0$. Indeed, if G has a fairly regular boundary γ, and ψ, which vanishes outside G, has continuous boundary values, then the equation (66) would not be true for $g = \operatorname{grad} \psi$ unless one adds at one side the term $\int \psi \cdot p_n \, do$ extending over the part of the boundary γ enclosed in T.

For obvious reasons this procedure fails for the second boundary problem. However, the idea again works handsomely when we replace the second by the third problem. We study a vector field v of class Γ and set $v = 0$ outside G; moreover,

$$\bar{v} = v \text{ in } T, \qquad = 0 \text{ outside } T.$$

We form the vector potential

$$V(x) = \int_{x'} \frac{\bar{v}(x')}{|x - x'|}$$

and the integral

$$J = \frac{1}{4\pi} \int (\operatorname{div} V \cdot \operatorname{div} v + \operatorname{rot} V \cdot \operatorname{rot} v).$$

By transformations similar to the foregoing we find the inequality

$$\int_T v^2 \leqq \frac{E}{4\pi} \cdot \int_G \{(\operatorname{rot} v)^2 + (\operatorname{div} v)^2\},$$

which is to be compared with Lemma 4.

Any pair (g, ψ) in $\mathfrak{G}^+$ is the limit of pairs

$$g_\nu = \operatorname{rot} v_\nu, \qquad \psi_\nu = \operatorname{div} v_\nu \qquad (v_\nu \text{ of class } \Gamma).$$

Application of our inequality to $v_\nu - v_\mu$ yields a vector field v vanishing outside G such that

$$\int_T (v - v_\nu)^2 \to 0 \quad \text{for } \nu \to \infty$$

in any cube T. This v satisfies the conditions

$$\int_T (v \cdot \operatorname{rot} q) = \int_T (g \cdot q), \qquad \int_T (v \cdot \operatorname{grad} \sigma) = -\int_T \psi \cdot \sigma$$

for any scalar σ with continuous derivatives in the closed T which vanishes along the boundary and any vector field q of the nature indicated by that letter. These integral equations are the equivalent for the differential equations

$$g = \operatorname{rot} v, \qquad \psi = \operatorname{div} v$$

together with the boundary condition $v = 0$.

8. **Concluding remarks.** In two-dimensional space the conditions (3) and (4) for irrotational and solenoidal fields, read

$$\int \left(f_1 \frac{\partial \psi}{\partial x_2} - f_2 \frac{\partial \psi}{\partial x_1} \right) = 0, \qquad \int \left(f_1 \frac{\partial \psi}{\partial x_1} + f_2 \frac{\partial \psi}{\partial x_2} \right) = 0 \qquad (\psi \text{ of class } \Gamma).$$

When one considers (f_1, f_2) as a covariant vector, these conditions are invariant under conformal transformations of the coördinates x_1, x_2. Hence the theory applies to arbitrary Riemann surfaces, i.e., to two-dimensional manifolds whose conformal structure is locally Euclidean. It yields, among others, this result: If φ is a periodic function with continuous derivatives on the universal Abelian covering manifold $\bar{G}$ of a compact Riemann surface G, then there exists a harmonic function η with the same periods. This statement is practically equivalent, and on the basis of well-known results due to de Rham, completely equivalent to the existence of an Abelian integral of first kind with prescribed periods of its real part. An alternative procedure for the construction of these integrals is the one followed by the author in his book on Riemann surfaces cited before. The method of orthogonal projection in Hilbert space is a pleasant variant of the Dirichlet principle of minimization. However, it ought to be said that our present method shows its full strength only by the general way in which it allows taking boundary conditions into account. The argument of the last section does not go through in two-dimensional space, unless G is bounded or at least such that the integral over G of $1/(1 + |x|^2)$ is finite.

The method is easily adaptable to the problems of two- and three-dimensional *elasticity theory.*

In n dimensions the conditions (4), (3) read:

$$\int \sum_i f_i \frac{\partial \psi}{\partial x_i} = 0,$$

$$\int \sum_{i,k} f_i \frac{\partial v_{ik}}{\partial x_k} = 0.$$

ψ is an arbitrary scalar of class Γ and v_{ik} an arbitrary skew-symmetric tensor of rank 2 and class Γ. It is not difficult to see how to formulate the corresponding conditions for Pfaffian forms of higher rank. Here we meet with Hodge's investigations mentioned above. He observes that in a $2k$-dimensional space the conditions for a Pfaffian form of rank k involve only the conformal, not the full metric, structure of the space, and thus the theory of these forms and their periods goes through in a "conformally flat" space (a space endowed with a Riemann metric for which the conformal curvature vanishes).

INSTITUTE FOR ADVANCED STUDY.

REAL INVERSION FORMULAS FOR LAPLACE INTEGRALS

By Harry Pollard

1. **Introduction.** In this paper we shall be concerned with real inversion formulas for the Laplace integrals

$$f(x) = \int_0^\infty e^{-xt}\, d\alpha(t), \tag{1}$$

$$f(x) = \int_0^\infty e^{-xt} \phi(t)\, dt. \tag{2}$$

In (1) $\alpha(t)$ is of bounded variation in $(0, R)$ for all positive R, and normalized, that is,

$$\alpha(0) = 0; \qquad \alpha(t) = \tfrac{1}{2}[\alpha(t+) + \alpha(t-)] \qquad (t > 0).$$

In (2) $\phi(t)$ is assumed Lebesgue integrable in $(0, R)$ for all positive R.

We shall obtain new inversion formulas for (1) and (2) in terms of

I. $f(x)$ and its derivatives in a neighborhood of infinity,

II. $f(x)$ and its differences in a neighborhood of infinity,

III. $f(x)$ and its derivatives on a discrete set of points near infinity.

In the special case of the integral (2) it will suffice to know the derivatives or differences of high enough order.

Case I has been treated by several writers, particularly D. V. Widder. Theorem 1.1, which forms the basis of our work, is an extension of results due to Widder and the present author [2].[1]

Case II can be handled by means of an operator employed by Widder to solve the Hausdorff moment problem ([6], pp. 174 and 178). The result is that if (2) converges, then[2]

$$\phi(t) = \lim_{k\to\infty} \frac{(n+k+1)!}{n!\,k!}(-1)^k \Delta_1^k f(n+1), \qquad n = \left[\frac{k}{e^t - 1}\right],$$

for almost all $t > 0$. There is a similar result for the integral (1). Our formulas differ somewhat from these.

Case III does not appear to have been treated previously.[3]

The main theorem, which we now state, concerns the integral (2). D will

Received August 1, 1940.

[1] Numbers in square brackets refer to the bibliography at the end of the paper.

[2] As usual, $[a]$ denotes the largest integer in a.

[3] For a summary of methods of inversion when other information concerning $f(x)$ is given see [1], p. 1.

be used to denote the Lebesgue set of the determining function $\phi(t)$. This set consists of almost all points $t > 0$ and includes the points of continuity of $\phi(t)$.

THEOREM 1.1. *Let $\{\theta_k(t)\}$ be a sequence of real functions satisfying*[4]

$$\theta_k = o(k) \qquad (k \to \infty) \tag{3}$$

for each $t > 0$. If (2) *converges, then*

$$\phi(t) = \lim_{k\to\infty} \left(\frac{k+\theta_k}{t}\right)^{k+1} \frac{(-1)^k}{k!} f^{(k)}\left(\frac{k+\theta_k}{t}\right) \tag{4}$$

for all t in D.

If $\phi(t+)$, $\phi(t-)$ both exist at a point $t > 0$, and if

$$\theta_k = o(k^{\frac{1}{2}}) \qquad (k \to \infty), \tag{5}$$

then

$$\frac{\phi(t+) + \phi(t-)}{2} = \lim_{k\to\infty} \left(\frac{k+\theta_k}{t}\right)^{k+1} \frac{(-1)^k}{k!} f^{(k)}\left(\frac{k+\theta_k}{t}\right). \tag{6}$$

Several particular cases of this result are known ([4], p. 772; [6], pp. 122, 126; [2], p. 422). We shall prove it in §3, where it will be shown to yield as corollaries the following new inversions of (2).

COROLLARY 1.11. *If* (2) *converges, then*

$$\phi(t) = \lim_{k\to\infty} \left[\frac{k}{t}\right]^{k+1} \frac{(-1)^k}{k!} f^{(k)}\left(\left[\frac{k}{t}\right]\right), \tag{7}$$

$$\phi(t) = \lim_{x\to\infty} x^{k+1} \frac{(-1)^k}{k!} f^{(k)}(x) \Big|_{k=[xt]} \tag{8}$$

for all t in D.

COROLLARY 1.12. *Let $\{x_n\}$ be any sequence approaching $+\infty$. If* (2) *converges, then*

$$\phi(t) = \lim_{n\to\infty} \frac{(-1)^k}{k!} x_n^{k+1} f^{(k)}(x_n) \Big|_{k=[x_n t]} \tag{9}$$

for all t in D.

2. **Lemmas.** We devote this section to the statement of several lemmas needed in the remainder of our work.

LEMMA 1. *If* (1) *converges, there exist $A > 0$, $g > 0$ such that*

$$|\alpha(u)| \leqq Ae^{gu} \qquad (u > 0).$$

This is a known result ([5], p. 703, Lemma 2).

[4] In this theorem we do not require that the o-conditions be satisfied uniformly with respect to t.

LEMMA 2. *If* (1) *converges, there exist* $N > 0$, $g > 0$ *such that*

$$|f^{(k)}(x)| \leqq \frac{Nk!}{x^k} \qquad (x > g).$$

The proofs of this and the next lemma are easily supplied by the reader.

LEMMA 3. *If* $\{\theta_k\}$ *is any sequence satisfying* (3), *then for* k *large enough*

$$\left(1 + \frac{\theta_k}{k}\right)^{k+1} < 2e^{\theta_k}.$$

LEMMA 4. *If* $\{\theta_k\}$ *is any sequence satisfying* (5), *then*

$$\frac{(k+\theta_k)^{k+1}}{k!}\int_0^1 e^{-(k+\theta_k)u}u^k\,du \to \tfrac{1}{2} \qquad (k \to \infty), \tag{10}$$

$$\frac{(k+\theta_k)^{k+1}}{k!}\int_1^\infty e^{-(k+\theta_k)u}u^k\,du \to \tfrac{1}{2} \qquad (k \to \infty).$$

It is obviously enough to prove (10). Since, by [6], p. 115,

$$\frac{k^{k+1}}{k!}\int_0^1 e^{-ku}u^k\,du \to \tfrac{1}{2} \qquad (k \to \infty),$$

it suffices to show that

$$J_k = \frac{k^{k+1}}{k!}\int^{1+\theta_k/k} e^{-ku}u^k\,du = o(1) \qquad (k \to \infty).$$

The maximum of $e^{-u}u$ occurs at $u = 1$, so that

$$|J_k| \leqq \frac{k^{k+1}}{k!}e^{-k}\frac{|\theta_k|}{k} \leqq |\theta_k|k^{-\frac{1}{2}},$$

the latter inequality being a consequence of Stirling's formula. An application of (5) completes the proof.

LEMMA 5. (10) *fails to hold when* $\theta_k = k^{\frac{1}{2}}$.

For by [3], vol. I, p. 80, Problem 210

$$\frac{(k+k^{\frac{1}{2}})^{k+1}}{k!}\int_0^1 e^{-(k+k^{\frac{1}{2}})u}u^k\,du = \frac{1}{k!}\int_0^{k+k^{\frac{1}{2}}} e^{-t}t^k\,dt$$

$$\to \tfrac{1}{2} + (2\pi)^{-\frac{1}{2}}\int_0^1 e^{-\frac{1}{2}t^2}\,dt \qquad (k \to \infty).$$

3. **The Laplace-Lebesgue integral.** In this section the integral (2) is treated, the general case (1) being left for §4.

We begin with a proof of the first part of Theorem 1.1. Let t be any point of D. Define

$$\beta(u) = \int_t^u \{\phi(v) - \phi(t)\}\,dv.$$

By the method of Widder ([6], pp. 122–124) the question is easily reduced to one of showing that

$$I_k = \frac{(k+\theta_k)^{k+1}}{(k-1)!}\int_0^\infty e^{-(k+\theta_k)u}\beta(tu)u^{k-1}\left(1 - \frac{k+\theta_k}{k}u\right)du = o(1) \qquad (k \to \infty).$$

Let $\epsilon > 0$ be arbitrary. Since t belongs to D, there exists a δ $(0 < \delta < 1)$ so small that

$$(11) \qquad | \beta(tu) | < \epsilon \, | u - 1 | \qquad (| u - 1 | < \delta).$$

Divide the interval of integration into the intervals $(0, 1-\delta)$, $(1-\delta, 1+\delta)$, $(1+\delta, \infty)$ and denote the corresponding integrals into which $I(k)$ is split by $I_1(k)$, $I_2(k)$, $I_3(k)$.

The functions $e^{-(k+\theta_k)u}u^{k-1}$, for large enough k, have their maxima at points greater than $1-\delta$ and are increasing in $(0, 1-\delta)$. Hence by (3) and Lemma 3

$$\begin{aligned} | I_1(k) | &\leqq \frac{(k+\theta_k)^{k+1}}{(k-1)!}\int_0^{1-\delta} e^{-(k+\theta_k)u}u^{k-1} | \beta(tu) | \left| 1 - \frac{k+\theta_k}{k}u \right| du \\ &\leqq A_1 \frac{k^{k+1}}{(k-1)!} e^{\theta_k} e^{-(k+\theta_k)(1-\delta)}(1-\delta)^{k-1}\int_0^1 | \beta(tu) | (1-u)\,du, \end{aligned}$$

where A_1 is a positive number independent of k. Since for k large enough $\theta_k \leqq \frac{1}{2}\delta k$, we have

$$| I_1(k) | \leqq A_2 \frac{k^{k+1}}{(k-1)!} e^{-k(1-\delta)} e^{\frac{1}{2}\delta^2 k}(1-\delta)^{k-1},$$

where A_2 is independent of k. By the test-ratio test the right side of this inequality is the general term of a convergent series, and hence approaches zero with $1/k$.

Similarly, $I_3(k)$ is seen to approach zero with $1/k$.

It remains to consider $I_2(k)$. By (11)

$$(12) \qquad | I_2(k) | < \epsilon \int_{1-\delta}^{1+\delta} | P_k(u) | \, du,$$

where

$$P_k(u) = \frac{(k+\theta_k)^{k+1}}{(k-1)!} e^{-(k+\theta_k)u}u^{k-1}(1-u)\left(1 - \frac{k+\theta_k}{k}u\right) \qquad (0 \leqq u < \infty).$$

k is to be taken so large that $k + \theta_k > 0$ for $k > k_0$. $P_k(u)$ changes sign at $u = 1$, $u = k/(k+\theta_k)$ and is positive outside the interval determined by these two points. Hence by (12) it follows that for $k > k_1 > k_0$

$$| I_2(k) | < \epsilon \int_{1-\delta}^{1+\delta} P_k(u)\,du + 2\epsilon \left| \int_{k/(k+\theta_k)}^{1} P_k(u)\,du \right|.$$

By (3) it is easily verified that for k large enough

$$\left| \int_{k/(k+\theta_k)}^{1} P_k(u)\, du \right| \leqq \int_0^{1-\delta} P_k(u)\, du,$$

$$\left| \int_{k/(k+\theta_k)}^{1} P_k(u)\, du \right| \leqq \int_{1+\delta}^{\infty} P_k(u)\, du,$$

so that by use of the gamma function

$$| I_2(k) | < \epsilon \int_0^{\infty} P_k(u)\, du = \epsilon.$$

This proves the first part of Theorem 1.1.

To prove the second part of the theorem define

$$w(u) = \phi(t-) \quad (0 < u < t); \qquad w(t) = \phi(t); \qquad w(u) = \phi(t+) \quad (u > t).$$

Let

$$F(x) = \int_0^{\infty} e^{-xu} w(u)\, du$$
$$= \phi(t-) \int_0^t e^{-xu}\, du + \phi(t+) \int_t^{\infty} e^{-xu}\, du.$$

Then by Lemma 4

$$\lim_{k\to\infty} \left(\frac{k+\theta_k}{t}\right)^{k+1} \frac{(-1)^k}{k!} F^{(k)}\left(\frac{k+\theta_k}{t}\right) = \frac{\phi(t+)}{2} + \frac{\phi(t-)}{2}. \tag{13}$$

Now let

$$p(u) = \phi(u) - w(u),$$
$$G(x) = \int_0^{\infty} e^{-xu} p(u)\, du.$$

$p(u)$ is continuous at $u = t$ and vanishes there. Then t belongs to the Lebesgue set of $p(u)$ and by the first part of the theorem

$$\lim_{k\to\infty} \left(\frac{k+\theta_k}{t}\right)^{k+1} \frac{(-1)^k}{k!} G^{(k)}\left(\frac{k+\theta_k}{t}\right) = 0. \tag{14}$$

Since $f(x) = F(x) + G(x)$, (6) follows from (13) and (14).

Condition (3) in the hypothesis of Theorem 1.1 cannot be improved if the conclusion (4) is to hold for *all* Laplace integrals; for let $f(x) = 1/x^2$. In particular cases, however, (3) may be needlessly restrictive, as, for example, if $f(x) = 1/x$.

Condition (5) of the same theorem cannot be replaced by $\theta_k = O(k^{\frac{1}{2}})$. For, by Lemma 5, (6) fails to hold at $t = 1$ when $\theta_k = k^{\frac{1}{2}}$ and $\phi(t)$ is defined by

$$\phi(t) = 1 \ (0 \leqq t < 1); \qquad \phi(1) = 1; \qquad \phi(t) = 0 \ (1 < t < \infty).$$

As for the corollaries of Theorem 1.1 stated in the introduction, (7) follows from (4) on letting

$$\left[\frac{k}{t}\right] = \frac{k + \theta_k(t)}{t};$$

(8) on letting

$$x = \frac{k + \theta_k(t)}{t},$$

where $k = [xt]$. Corollary 1.12 is an immediate consequence of Corollary 1.11.

Note that (4) and (8) furnish inversion formulas for Case I mentioned in the introduction, while (7) and (9) cover Case III.

To complete the remaining case we introduce a linear difference operator. Let $f(x)$ be defined by a Laplace integral for $x > \sigma_c$. Then for any $t > 0$ we define

$$M_{k,t}^{\lambda_k}\{f(x)\} = \frac{(-1)^k}{k!} x^{k+1} \lambda_k^{-k} \Delta_{\lambda_k}^k f(x) \Big|_{x=k/t}, \tag{15}$$

where $\{\lambda_k\}$ is a sequence such that

$$\frac{k}{t} + k\lambda_k > \sigma_c \qquad (k = 1, 2, \cdots) \tag{16}$$

and

$$\Delta_{\lambda_k}^k f(x) = \sum_{n=0}^{k} \binom{k}{n} (-1)^{k-n} f(x + n\lambda_k).$$

The condition (16) guarantees that (15) is well-defined.

We conclude this section by showing how the operator (15) serves to invert the Laplace integral.

THEOREM 3.1. *Let* (2) *converge. Then*

$$\phi(t) = \lim_{k\to\infty} M_{k,t}^{\lambda_k}\{f(x)\}$$

for all t in D, if

$$\lambda_k = o\left(\frac{1}{k}\right) \qquad (k \to \infty). \tag{17}$$

By a well-known generalization of Rolle's theorem

$$M_{k,t}^{\lambda_k}\{f(x)\} = \frac{(-1)^k}{k!}\left(\frac{k}{t}\right)^{k+1} f^{(k)}\left(\frac{k}{t} + \theta_k\lambda_k\right) \tag{18}$$

where $0 \leqq \theta_k \leqq k$. By (17), (18), and the first part of Theorem 1.1 it is enough to show that

$$k^{k+1} \sim (k + \theta_k\lambda_k t)^{k+1} \qquad (k \to \infty).$$

But by (17)

$$\left(\frac{k+\theta_k\lambda_k t}{k}\right)^{k+1} \sim e^{\theta_k\lambda_k t} \sim 1 \qquad (k \to \infty).$$

The proof is now complete.

4. **The Laplace-Stieltjes integral.** To obtain the corresponding inversions for (1) we begin by observing that we can write

$$f(x) = x\int_0^\infty e^{-xt}\alpha(t)\,dt$$

for sufficiently large values of x. This is a consequence of Lemma 1. Applying the second part of Theorem 1.1 to

$$F(x) = \frac{f(x)}{x} = \int_0^\infty e^{-xt}\alpha(t)\,dt,$$

we deduce by Leibniz' rule

THEOREM 4.1. *If* (1) *converges, then*

$$\alpha(t) = \lim_{k\to\infty}\sum_{n=0}^{k}\frac{(-1)^n}{n!}\left(\frac{k+\theta_k}{t}\right)^n f^{(k)}\left(\frac{k+\theta_k}{t}\right) \qquad (t > 0),$$

the functions $\{\theta_k(t)\}$ *satisfying* (5) *for each* $t > 0$.[4]

The case $\theta_k = 0$ $(k = 1, 2, \cdots)$ is due to Widder, the case $0 \leqq \theta_k \leqq 1$ to the author [2].

As above we deduce immediately the following corollaries.

COROLLARY 2.11. *If* (1) *converges, then*

$$\alpha(t) = \lim_{k\to\infty}\sum_{n=0}^{k}\frac{(-1)^n}{n!}\left[\frac{k}{t}\right]^n f^{(n)}\left(\left[\frac{k}{t}\right]\right) \qquad (t > 0),$$

$$\alpha(t) = \lim_{x\to\infty}\sum_{n=0}^{[xt]}\frac{(-1)^n}{n!}x^n f^{(n)}(x) \qquad (t > 0).$$

The second of these is the Feller-Dubourdieu inversion of (1) ([2], p. 420).

COROLLARY 2.12. *Let* $\{x_m\}$ *be any sequence approaching* $+\infty$. *If* (1) *converges, then*

$$\alpha(t) = \lim_{m\to\infty}\sum_{n=0}^{[x_m t]}\frac{(-1)^n}{n!}x_m^n f^{(n)}(x_m) \qquad (t > 0).$$

To obtain the inversion which corresponds to Theorem 3.1 we can apply the result given there to the function $F(x) = f(x)/x$. We prefer, however, the following form.

THEOREM 4.2. *Let* (1) *converge. Then*

$$\alpha(t) - f(\infty) = \lim_{k\to\infty}\int_0^t M_{k,u}^{\lambda_k}\{f(x)\}\,du \qquad (t > 0),$$

if

$$\lambda_k = o(1/k^2) \qquad (k \to \infty). \tag{19}$$

By a theorem of Widder ([7], p. 249)

$$\alpha(t) - f(\infty) = \lim_{k\to\infty} \int_0^t \frac{(-1)^k}{k!} \left(\frac{k}{u}\right)^{k+1} f^{(k)}\left(\frac{k}{u}\right) du \qquad (t > 0).$$

Hence by (18) it suffices to show that for any fixed $t > 0$

$$H_k = \int_0^t \frac{(-1)^k}{k!} \left(\frac{k}{u}\right)^{k+1} \left\{ f^{(k)}\left(\frac{k}{u} + \theta_k\lambda_k\right) - f^{(k)}\left(\frac{k}{u}\right)\right\} du$$

$$= \frac{(-1)^k}{(k-1)!} \int_{k/t} u^{k-1} \{f^{(k)}(u + \phi_k\lambda_k) - f^{(k)}(u)\}\, du = o(1) \qquad (k \to \infty),$$

where $0 \leqq \phi_k(u) \leqq k$. By Rolle's theorem

$$| H_k | \leqq \frac{1}{(k-1)!} \int_{k/t}^{\infty} u^{k-1} | \phi_k\lambda_k | \, | f^{(k+1)}(u + \psi_k\lambda_k) | \, du,$$

where $0 \leqq \psi_k(u) \leqq k$. Then by Lemma 2 and (19) we have for k large enough

$$| H_k | \leqq \frac{k\,|\lambda_k|}{(k-1)!} \int_{k/t}^{\infty} u^{k-1} N \frac{(k+1)!}{u^{k+1}} du = N\,|\lambda_k|\, tk(k+1) = o(1) \qquad (k \to \infty).$$

References

1. R. P. Boas, Jr., and D. V. Widder, *An inversion formula for the Laplace integral*, this Journal, vol. 6(1940), pp. 1–26.
2. H. Pollard, *Note on the inversion of the Laplace integral*, this Journal, vol. 6(1940), pp. 420–424.
3. G. Pólya and G. Szegö, *Aufgaben und Lehrsätze aus der Analysis*, Berlin, 1925.
4. E. L. Post, *Generalized differentiation*, Transactions of the American Mathematical Society, vol. 32(1930), pp. 723–781.
5. D. V. Widder, *A generalization of Dirichlet's series and of Laplace's integrals by means of a Stieltjes integral*, Transactions of the American Mathematical Society, vol. 31(1929), pp. 694–743.
6. D. V. Widder, *The inversion of the Laplace integral and the related moment problem*, Transactions of the American Mathematical Society, vol. 36(1934), pp. 107–200.
7. D. V. Widder, *A classification of generating functions*, Transactions of the American Mathematical Society, vol. 39(1936), pp. 244–298.

Harvard University.

CHARACTERISTIC FUNCTIONS OF FAMILIES OF SETS

By M. H. Stone

In an interesting paper entitled *The characteristic function of a sequence of sets and some of its applications*, Fundamenta Mathematicae, vol. 31(1938), pp. 207–233 (see also Fundamenta Mathematicae, vol. 26(1935), p. 302) Szpilrajn has employed the characteristic function to develop a certain method of dealing with the algebraic structure of sequences of sets; and has established with the aid of this method a variety of specific theorems and equivalences in the domain of set-theoretical topology. He attributes to Kuratowski the first use of the characteristic function of a sequence of sets.

In the present note, I shall trace certain connections between the content of Szpilrajn's paper and the general theory of abstract Boolean algebras which I have developed in two memoirs published elsewhere: *The theory of representations for Boolean algebras*, Transactions of the American Mathematical Society, vol. 40(1936), pp. 37–111 (cited here by the letter R); and *Applications of the theory of Boolean rings to general topology*, ibid., vol. 41(1937), pp. 375–481 (cited here by the letter A). In doing so, I deem my chief purpose to be that of reconciling two independent points of view which prove, upon examination, to present a considerable similarity so far as the theory of the algebraic structure of sequences of sets is concerned.

As I shall point out below, an obvious but theoretically desirable generalization of Szpilrajn's work leads to the introduction of the characteristic function of an arbitrary transfinite sequence, or well-ordered family, of sets. It seems to me of more importance, perhaps, to observe that the rôle of order, which is essential to the definition of the characteristic function, appears to be artificial so far as the majority of applications is concerned. In principle, therefore, one is tempted to seek an order-free theory of the algebraic relations envisaged. I shall show here that such a theory is already in existence and that, through the adjunction of elementary considerations of order, it leads back to the theory of the characteristic function due to Szpilrajn.

1. The space $\mathfrak{B}_\mathfrak{c}$. If $\mathfrak{c}$ is any cardinal number, we shall denote by $\mathfrak{B}_\mathfrak{c}$ the Cartesian product of $\mathfrak{c}$ two-point Hausdorff spaces. It is a totally-disconnected bicompact Hausdorff space; in the particular case where $\mathfrak{c} = \aleph_0$ it is homeomorphic with the Cantor discontinuum. In the sequel we shall suppose that $\mathfrak{c}$ is an infinite cardinal.

Received August 10, 1940. This paper was originally accepted for publication in volume 33 of Fundamenta Mathematicae. Page proof was read in the early summer of 1939. Presumably that volume will not appear in the immediate future.

A more detailed description of $\mathfrak{B}_c$, adapted to later discussions, may be given in the following terms. Starting from the class A of ordinals α preceding some fixed ordinal β, subject to the restriction that A shall have cardinal number c, we take as the points of $\mathfrak{B}_c$ the real functions $\mathfrak{s}$ defined over A and assuming only the values 0, 1. We then take as a basis (of open sets) for $\mathfrak{B}_c$ the family composed of all sets $\mathfrak{E}_\alpha$ and $\mathfrak{E}'_\alpha$, $\alpha < \beta$, where $\mathfrak{E}_\alpha$ consists of all $\mathfrak{s}$ with $\mathfrak{s}(\alpha) = 1$ and $\mathfrak{E}'_\alpha$ is the complement of $\mathfrak{E}_\alpha$ in $\mathfrak{B}_c$ (i.e., $\mathfrak{E}'_\alpha$ consists of all $\mathfrak{s}$ with $\mathfrak{s}(\alpha) = 0$). The algebraic and topological properties of the space $\mathfrak{B}_c$ so obtained are fully discussed in R Chapter I, especially Theorems 9-13; they include those noted above. For our later convenience we have replaced the sets $\mathfrak{U}_\alpha$, $\mathfrak{U}'_\alpha$ of R by $\mathfrak{E}'_\alpha$, $\mathfrak{E}_\alpha$ respectively in the present description.

The sets $\mathfrak{E}_\alpha$ bear a special relationship to the topology of $\mathfrak{B}_c$ which proves to be of importance in the sequel. This relationship is most concisely expressed in terms of the Boolean algebra A_c generated from the sets $\mathfrak{E}_\alpha$ by the formation of all possible complements, finite unions, and finite intersections. According to R Definitions 1, 2 and Theorems 1, 9 the sets belonging to A_c are characterized as the open-and-closed subsets of $\mathfrak{B}_c$. Moreover, R Theorem 12 discloses that the algebra A_c is a free Boolean algebra generated by the c elements $\mathfrak{E}_\alpha$ and is thus completely characterized in algebraic terms. The proof of the theorem cited consists partly in establishing the independence (to use Szpilrajn's terminology) of the sequence $\{\mathfrak{U}_\alpha\} = \{\mathfrak{E}'_\alpha\}$, which is obviously equivalent to that of $\{\mathfrak{E}_\alpha\}$. The constituents of $\{\mathfrak{E}_\alpha\}$ are thus non-void closed subsets of $\mathfrak{B}_c$. By virtue of the bicompactness of $\mathfrak{B}_c$, all the atoms of $\{\mathfrak{E}_\alpha\}$ are therefore non-void; and, since the sets $\mathfrak{E}_\alpha$, $\mathfrak{E}'_\alpha$, $\alpha < \beta$, constitute an open basis for $\mathfrak{B}_c$, no atom can contain more than one point. The sequence $\{\mathfrak{E}_\alpha\}$ is therefore completely independent. As a matter of fact it is quite easy to see, directly from the definition of the sets $\mathfrak{E}_\alpha$ and without recourse to the topology of $\mathfrak{B}_c$, not only that the constituents of $\{\mathfrak{E}_\alpha\}$ are non-void, the sequence consequently being independent, but also that the atoms of $\{\mathfrak{E}_\alpha\}$ are precisely the one-point subsets of $\mathfrak{B}_c$, the sequence thus being completely independent. And the bicompactness of $\mathfrak{B}_c$ can then be deduced essentially from this direct observation in the manner indicated in the proof of R Theorem 9.

2. Definition of the characteristic function. Let $e = \{E_\alpha\}$, where α ranges over the class A of §1, be a transfinite sequence of subsets of a fixed non-void set X. The characteristic function c_e of this sequence is defined as that single-valued function from X into $\mathfrak{B}_c$ determined by assigning to each x in X the image $\mathfrak{s} = c_e(x)$, where $\mathfrak{s}(\alpha)$ is 0 or 1 according as $x \in E'_\alpha$ or $x \in E_\alpha$, $\alpha < \beta$. It is evident that an arbitrary single-valued function c from X into $\mathfrak{B}_c$ is the characteristic function of a unique sequence $e = \{E_\alpha\}$ determined by taking E_α as the set of all x such that $\mathfrak{s} = c(x)$ satisfies the condition $\mathfrak{s}(\alpha) = 1$. In short, the characteristic function c_e is determined by the relations $c_e(E_\alpha) = \mathfrak{E}_\alpha c_e(X)$, $E_\alpha = c_e^{-1}(\mathfrak{E}_\alpha)$, $X = c_e^{-1}c_e(X)$.

These relations evidently serve to establish the two following properties,

which summarize the information essential to the majority of applications of the characteristic function:

(1) the correspondence $\mathfrak{E}_\alpha c_e(X) \leftrightarrow E_\alpha$ induces an isomorphism between the Boolean algebra $A_\mathfrak{c}(X)$ generated in $c_e(X)$ by the sets $\mathfrak{E}_\alpha c_e(X)$ and the Boolean algebra $\mathbf{E}_0$ generated in X by the sets E_α;

(2) the sets in $A_\mathfrak{c}(X)$ are both open and closed in the relative topology of $c_e(X)$ considered as a subspace of $\mathfrak{B}_\mathfrak{c}$.

From A Theorem 56 we know that the correspondence $\mathfrak{E}_\alpha \to \mathfrak{E}_\alpha c_e(X)$ induces a homomorphism $A_\mathfrak{c} \to A_\mathfrak{c}(X)$. Hence the correspondence $\mathfrak{E}_\alpha \leftrightarrow E_\alpha$ induces a homomorphism $A_\mathfrak{c} \to \mathbf{E}_0$.

It is convenient to determine when two elements x_1 and x_2 in X satisfy the condition $c_e(x_1) = c_e(x_2)$. Without difficulty we find that the following statements are equivalent: $c_e(x_1) = c_e(x_2)$; x_1 and x_2 are separated by $\{E_\alpha\}$; x_1 and x_2 are separated by $\mathbf{E}_0$; x_1 and x_2 belong to the same atom of $\{E_\alpha\}$. Consequently we may construct c_e in two steps, first identifying points of X belonging to the same atom of $\{E_\alpha\}$ and then forming the characteristic function of the reduced sequence $\{E_\alpha^*\}$ in the resulting set X^*. It is evident that the first step here is precisely that of reducing the algebra $\mathbf{E}_0$ in accordance with A Theorem 54.

3. Alternative definition of the characteristic function. We shall now rephrase the definition of the characteristic function in algebraic terms.

We begin by observing that the correspondence $\mathfrak{E}_\alpha \leftrightarrow E_\alpha$, $\alpha < \beta$, induces a homomorphism $A_\mathfrak{c} \to \mathbf{E}_0$. Since $A_\mathfrak{c}$ is a free Boolean algebra, each of its elements is expressible as a finite union of terms of the general form

$$\mathfrak{E}_{\alpha_1} \cdots \mathfrak{E}_{\alpha_n}\mathfrak{E}'_{\alpha_{n+1}} \cdots \mathfrak{E}'_{\alpha_{n+p}} \qquad (n \geqq 0, p \geqq 0, n + p \geqq 1):$$

and, as is shown in the proof of R Theorem 12, two such expressions can be equal *only* in consequence of the fundamental Boolean identities. On replacing each $\mathfrak{E}_\alpha$ in every such expression by its correspondent E_α, we obtain a correspondence from $A_\mathfrak{c}$ to $\mathbf{E}_0$ which carries equal elements into equal elements, complements into complements, unions into unions, and intersections into intersections, by virtue of the fundamental Boolean identities. We thus have a homomorphism $A_\mathfrak{c} \to \mathbf{E}_0$, which evidently becomes an isomorphism if and only if $\{E_\alpha\}$ is independent.

With each point x in X we now associate, by virtue of A Theorem 34, the prime ideal $\mathfrak{p}(x)$ in $\mathbf{E}_0$ which consists of those sets in $\mathbf{E}_0$ *not* containing x. From $\mathfrak{p}(x)$ we pass by A Theorem 48 to the prime ideal $\mathfrak{p}_\mathfrak{c}(x)$ of all those sets in $A_\mathfrak{c}$ which are carried by the homomorphism $A_\mathfrak{c} \to \mathbf{E}_0$ into sets in $\mathfrak{p}(x)$. And from $\mathfrak{p}_\mathfrak{c}(x)$ we pass to a uniquely determined point $\mathfrak{s}_x$ in $\mathfrak{B}_\mathfrak{c}$ with the help of R Theorem 9.

The characteristic function c_e of $\{E_\alpha\}$ is defined by the equation $c_e(x) = \mathfrak{s}_x$ for all x in X.

It is easy to show this definition equivalent to that of §2. The proof of R Theorem 9 discloses that $\mathfrak{s}_x(\alpha)$ is 0 or 1 according as $\mathfrak{U}'_\alpha = \mathfrak{E}_\alpha \in \mathfrak{p}_\mathfrak{c}(x)$ or $\mathfrak{U}_\alpha = \mathfrak{E}'_\alpha \in \mathfrak{p}_\mathfrak{c}(x)$. Clearly the relations $\mathfrak{E}_\alpha \in \mathfrak{p}_\mathfrak{c}(x)$, $\mathfrak{E}'_\alpha \in \mathfrak{p}_\mathfrak{c}(x)$ are equivalent respec-

tively to the relations $E_\alpha \in \mathfrak{p}(x)$, $E'_\alpha \in \mathfrak{p}(x)$; and hence are equivalent respectively to the relations $x \in E'_\alpha$, $x \in E_\alpha$. Thus we see that $c_e(x) = \mathfrak{s}_x$ is determined by putting $\mathfrak{s}_x(\alpha)$ equal to 0 or to 1 according as $x \in E'_\alpha$ or $x \in E_\alpha$.

4. **Unordered families.** The construction of the characteristic function c_e described in the preceding section reveals that the ordering of $\{E_\alpha\}$ has significance only insofar as it determines the homomorphism $A_c \to \mathbf{E}_0$ through the correspondence $\mathfrak{E}_\alpha \leftrightarrow E_\alpha$. If we replace this homomorphism by any other, the construction can still be carried through and still provides a mapping of X in $\mathfrak{B}_c$ with the essential properties (1) and (2) of §2. By considering the reduced algebra determined by $\mathbf{E}_0$, as suggested at the close of §2, and applying A Theorem 69 we obtain an immediate verification of property (1); and property (2) then follows from the results summarized in §1.

The most direct treatment of unordered families on the basis of A and R is, however, the following. Let $\{E\}$ be a family of subsets, distinct or not, of a fixed non-void set X; and let $\mathbf{E}_0$ be the Boolean algebra generated in X by the given sets E. By A Theorem 67 and R Theorem 1 there is associated with $\mathbf{E}_0$ a totally-disconnected bicompact Hausdorff space $\mathfrak{B}(\mathbf{E}_0)$. If the algebra $\mathbf{E}_0$ is reduced in accordance with A Theorem 54, the resulting isomorphic algebra of sets is, by virtue of A Theorem 69, equivalent to a certain algebra of subsets of a fixed set $\mathfrak{X} \subset \mathfrak{B}(\mathbf{E}_0)$; and, in view of information summarized in A Theorem 69 and R Theorem 1, the latter algebra consists precisely of the sets $\mathfrak{G}\mathfrak{X}$, where $\mathfrak{G}$ is both open and closed in $\mathfrak{B}(\mathbf{E}_0)$, the set $\mathfrak{X}$ itself being everywhere dense in $\mathfrak{B}(\mathbf{E}_0)$. Thus we obtain a single-valued function c from X onto $\mathfrak{X}$ in $\mathfrak{B}(\mathbf{E}_0)$; and we see that c carries $\mathbf{E}_0$ isomorphically into the algebra of sets $\mathfrak{G}\mathfrak{X}$, the inverse c^{-1} serving to invert the isomorphism determined by c. If we now appeal to R Theorem 10, we can imbed $\mathfrak{B}(\mathbf{E}_0)$ topologically, as a closed set, in the space $\mathfrak{B}_c$. The function c therefore maps X into $\mathfrak{B}_c$, with $c(X) = \mathfrak{X}$ as before; and the algebra into which it carries $\mathbf{E}_0$ can now be characterized as consisting of all sets $\mathfrak{G}_c\mathfrak{X}$, where $\mathfrak{G}_c$ is both open and closed in $\mathfrak{B}_c$ (i.e., is an element of A_c).

5. **An application.** To illustrate the applicability of the results of §4 we shall consider a generalization of a theorem of Kuratowski proved by Szpilrajn with the aid of the characteristic function. Let X be a non-void topological space—more specifically, a T_0-space (in the terminology of Alexandroff and Hopf) of infinite character $\mathfrak{c}$. Let $\{E\}$ be an arbitrary (open) basis for X. We may in particular choose $\{E\}$ so that its cardinal number is $\mathfrak{c}$. Since any two distinct points in X are separated by the basis $\{E\}$, the Boolean algebra $\mathbf{E}_0$ is a reduced algebra of sets. Thus the map c of X onto $\mathfrak{X} \subset \mathfrak{B}_c$ is biunivocal. Now, if G is any non-void open subset of X, it is the union of certain sets E; and $c(G)$ is the union of the corresponding sets $c(E)$. Since $c(E)$ is open in the relative topology of $\mathfrak{X}$, considered as a subspace of $\mathfrak{B}_c$, we conclude that $c(G)$ is also open in this topology. We thus obtain the following result:

THEOREM. *If X is any T_0-space of infinite character $\mathfrak{c}$, it is a biunivocal continuous image (by the map c^{-1}) of an appropriate subspace $\mathfrak{X}$ of a totally-disconnected bicompact Hausdorff space of character $\mathfrak{c}$ (the space $\mathfrak{B}_\mathfrak{c}$).*

Of course, the map c of X onto $\mathfrak{X}$ is continuous only if X is homeomorphic with $\mathfrak{X}$. The conditions under which a given space X is homeomorphic with a subspace of a totally-disconnected bicompact Hausdorff space (Boolean space) are discussed in R Theorem 55. Even though c fails to be continuous, its topological character must still be comparatively simple. In fact we can show that, if $\mathfrak{G}$ is any (relatively) open subset of $\mathfrak{X}$, then $c^{-1}(\mathfrak{G})$ is the union of at most $\mathfrak{c}$ sets of the form GF, where G is open and F closed in X. We include here the cases $G = X, F = X$. From §4 we know that the sets $c(H)$, $H \epsilon \mathrm{E}_0$, constitute a basis for $\mathfrak{X}$. Hence $\mathfrak{G}$ is the union of (at most $\mathfrak{c}$) sets $c(H)$; and $c^{-1}(\mathfrak{G})$ is the union of the corresponding sets $H = c^{-1}c(H)$. Now each set H, being expressible in terms of the basis $\{E\}$ as a finite union of sets of the form

$$E_1 \cdots E_n E'_{n+1} \cdots E'_{n+p} \qquad (n \geqq 0, p \geqq 0, n + p \geqq 1),$$

is a finite union of sets of the form GF described above. Thus $c^{-1}(\mathfrak{G})$ has the property asserted. In case X is a regular space we can sharpen the preceding result, asserting now that $c^{-1}(\mathfrak{G})$ is the union of at most $\mathfrak{c}$ closed sets. To establish this proposition we associate with each point x of a non-void open set $G \subset X$ an open set $G(x)$ such that $x \epsilon G(x) \subset G^{-}(x) \subset G$, this being possible by the regularity of X. Since $\{E\}$ is a basis for X, it must contain a set $E(x)$ such that $x \epsilon E(x) \subset G(x)$. The relations $E^{-}(x) \subset G^{-}(x) \subset G$ now show that G is the union of the closed sets $E^{-}(x)$, of which at most $\mathfrak{c}$ are distinct. It follows immediately that $c^{-1}(\mathfrak{G})$, being the union of at most $\mathfrak{c}$ sets GF, can be expressed as the union of at most $\mathfrak{c}$ closed sets. In particular, if X is regular and separable —in other words, is metric and separable—, the set $c^{-1}(\mathfrak{G})$ is an F_σ-set.

In conclusion we point out that the representation provided by the theorem proved above is entirely distinct from the representations provided in the theory of "Boolean maps" given in R. Indeed the present representation of X in $\mathfrak{B}_\mathfrak{c}$ defines a Boolean map $m(\mathfrak{B}_\mathfrak{c}, \mathfrak{X}^*, X)$, where $\mathfrak{X}^*$ is the family of one-point subsets of $\mathfrak{X} = c(X)$, only in the extremely special case where X and $\mathfrak{X}$ are homeomorphic by the map c.

HARVARD UNIVERSITY.

A CLASS OF DIFFERENTIAL OPERATORS OF INFINITE ORDER, I

By Einar Hille

Introduction. The present paper is the first part of an investigation devoted to the theory of differential operators of infinite order[1] of the form

$$G(\delta_z) = \sum_{k=0}^{\infty} g_k \delta_z^k. \tag{1}$$

Here

$$G(w) = \sum_{k=0}^{\infty} g_k w^k$$

is supposed to be an entire function,[2] the order and type of which will be subjected to various restrictions;

$$\delta_z = z^2 - \frac{d^2}{dz^2} \tag{2}$$

is the differential operator of Hermite-Weber; and $\delta_z^k = \delta_z \cdot \delta_z^{k-1}$. Putting

$$h_n(z) = (-1)^n e^{\frac{1}{2}z^2} \frac{d^n}{dz^n}(e^{-z^2}) \equiv e^{-\frac{1}{2}z^2} H_n(z), \tag{3}$$

where $H_n(z)$ is the n-th polynomial of Hermite, we find that

$$\delta_z h_n(z) = (2n + 1) h_n(z). \tag{4}$$

The author has shown the importance of the differential operator $G(\delta_z)$ in the theory of Hermite series (see E. Hille [4]). There only those features of the theory were discussed which were of immediate use for Hermite series. In the present paper and its continuation we shall consider various questions omitted in the earlier discussion.

The basic notion of applicability of a differential operator was given on page 897 of the paper quoted above. Let $G(w)$ be a given entire function and let $\mathfrak{F}$ be a given class of analytic functions $\{f(z)\}$. We say that *the differential operator $G(\delta_z)$ applies to or is applicable to the class $\mathfrak{F}$ if the series*

$$G(\delta_z) \cdot f(z) = \sum_{k=0}^{\infty} g_k \delta_z^k f(z) \tag{5}$$

Received September 16, 1940; presented to the American Mathematical Society, September 10, 1940.

[1] For a survey of the general field of differential operators of infinite order see R. D. Carmichael [1] and H. T. Davis [2]. The latter has an extensive bibliography. Numbers in brackets refer to the bibliography at the end of this paper.

[2] For the theory of entire functions used in this paper consult the treatise of G. Valiron [9].

converges at every point where $f(z)$ is holomorphic, the sum of the series being a holomorphic function of z in any domain in which $f(z)$ is holomorphic, no matter what element $f(z)$ of $\mathfrak{F}$ we substitute in the series.

It was shown on page 898 of the same paper that if $\mathfrak{F}$ is the class of all analytic functions, a necessary and sufficient condition in order that $G(\delta_z)$ shall apply to $\mathfrak{F}$ is that the entire function $G(w)$ be of order $\sigma \leqq \frac{1}{2}$ and of *minimal type* if $\sigma = \frac{1}{2}$. A new proof of this theorem will be given in §4 below; and in part II of this paper we shall show that the condition is still sufficient if we replace δ_z by an arbitrary second-order differential operator, the coefficients of which are entire functions. The result also extends to differential operators of order n if we replace $\frac{1}{2}$ by n^{-1}.

The present paper is devoted mainly to the applicability of the differential operator $G(\delta_z)$ to classes of entire functions. §§1–3 contain preliminary material such as formal representations of the operators, fundamental estimates of $\delta_z^k f(z)$ for large values of k under different assumptions on $f(z)$, and a brief discussion of the continuity properties of the functional

$$F_\sigma(z;f) = \limsup_{k\to\infty} \left| \frac{\delta_z^k f(z)}{\Gamma(1+k/\sigma)} \right|^{1/(2k)}, \tag{6}$$

which is analogous to the functional

$$\Phi_\sigma(z;f) = \limsup_{k\to\infty} \left| \frac{f^{(k)}(z)}{\Gamma(1+k/\sigma)} \right|^{1/k} \tag{7}$$

in the theory of the differential operator $G(d/dz)$. Most of the material in these sections is new either in form or in substance, but there is a certain amount of unavoidable overlapping with §3.3 of the previous paper. Whenever possible we refer the reader to this paper for further details, however.

The main theme occupies §§4–8. Certain results bearing on this problem ([4], Theorems 3.3 and 3.4) have already been announced without proofs. These results are restated in amplified form and proved in the present paper. The fundamental notions are the *order relation*, defining the *conjugate order*, the *conjugate type* and the *critical ρ-order*, the conjugate of which is the *maximal σ-order*. The order relation is

$$\frac{1}{\rho} + \frac{1}{2\sigma} = 1, \tag{8}$$

valid for $\rho \geqq 2$ which is the critical ρ-order with $\sigma = 1$ as the corresponding maximal σ-order. The order relation can be made more precise by the introduction of the conjugate type. Let it be known that[3]

$$\limsup_{r\to\infty} r^{-\rho} \log M(r;f) \leqq \alpha, \qquad \rho > 2, \tag{9}$$

[3] $M(r;f)$ denotes the maximum modulus of $f(z)$ on the circle $|z| = r$. Similar notation is used for other functions of z or w.

and let β be the conjugate type of α determined by the equation

$$(\rho\alpha)^{1/\rho}(2\sigma\beta)^{1/(2\sigma)} = 1, \tag{10}$$

where σ is given by (8); then $G(\delta_z)$ applies to $f(z)$ whenever

$$\limsup_{r\to\infty} r^{-\sigma} \log M(r; G) < \beta, \tag{11}$$

and this limit is the best of its kind. If $\rho = 2$ there is still a conjugate type depending upon α, though we cannot determine its exact value. For $\rho < 2$ there exists a maximal type $\beta(0)$ such that $G(\delta_z)$ applies whenever (11) holds with σ replaced by 1 and β by $\beta(0)$. Again, we can determine fairly narrow limits for $\beta(0)$ but not its exact value. We also investigate how the order and type of the transform $G(\delta_z)\cdot f(z)$ depend upon those of $f(z)$ and $G(w)$. To prove that the various limits obtained are the best possible and to prove the existence of various phenomena, we need a large number of counter examples which are assembled in §8.

In §9 we have joined various loosely connected remarks on $G(\delta_z)$ and related operators. The applicability theory extends to operators which are entire functions in δ_z with coefficients which are polynomials in z of limited degree. The greater part of the section contains general reflections on the outstanding differences between the theories of the operators $G(\delta_z)$ and $G(d/dz)$. The presence of a finite maximal σ-order is such a difference, the irregular behavior of the functional $F_\sigma(z, f)$, defined by (6), is another. The fact that the functions $f(z)$ for which $F_\infty(z, f)$ is bounded are highly specialized compared with the functions for which $\Phi_\infty(z, f)$ is bounded is also noteworthy. This implies that the classical method of solving non-homogeneous differential equations of infinite order by operator series or their equivalents is of very limited interest in case of the equation $G(\delta_z)\cdot W = F(z)$.[4] Finally, we call attention to the fact that the operator $G(\delta_z)$ seems to be oriented in the complex plane, and that the values of $G(w)$ on the fixed sets $\{2n + 1\}$ and $\{-2n - 1\}$ are of fundamental importance for the behavior of the operator. There is nothing corresponding to this situation in the theory of the operator $G(d/dz)$.

In later papers we shall study the equation

$$G(\delta_z)\cdot W = F(z) \tag{12}$$

for given functions $F(z)$ and $G(w)$. We shall also extend parts of the applicability theory to operators of the form $G(D_z)$, where D_z is a given second-order differential operator.

1. Basic formulas. In the following $f(z)$ is an analytic function holomorphic within a domain D. We shall obtain various expressions for the operators

[4] A postulational treatment of operational equations which also applies to the operator δ_z is due to F. Schürer [8].

$\delta_z^k f(z)$ with the aid of Cauchy's formulas. For z in D we have obviously

$$\delta_z f(z) = \frac{1}{2\pi i} \oint \frac{z^2(t-z)^2 - 2}{(t-z)^3} f(t)\, dt.$$

Here we can replace z^2 by t^2 in the numerator without changing the value of the integral. As contour of integration we can choose a small circle $| t - z | = R$. By iteration we get

$$(1.1) \qquad \delta_z^k f(z) = (2\pi i)^{-k} \oint \underset{(k)}{\cdots} \oint \left\{ \prod_{\nu=1}^{k} \frac{t_{\nu+1}^2 (t_\nu - t_{\nu+1})^2 - 2}{(t_\nu - t_{\nu+1})^3} \right\} f(t_1)\, dt_1 \cdots dt_k,$$

where t_{k+1} is to be replaced by z. The contours of integration can be taken as sufficiently small circles $| t_\nu - t_{\nu+1} | = R_\nu$, the sum of the radii of which is less than the distance from z to the boundary of D. The integrations are to be carried out in the natural order of the subscripts.

This formula is highly condensed and can be used for estimates but does not readily yield the best possible appraisals. In order to get fairly sharp estimates of the behavior of $\delta_z^k f(z)$ as $k \to \infty$ we had better resolve $\delta_z^k f(z)$ into components. The expression on the right side of (1.1) can evidently be written as the sum of the 2^k integrals

$$(1.2) \qquad I_{i_1 \cdots i_k} = (-2)^m (2\pi i)^{-k} \oint \underset{(k)}{\cdots} \oint \left\{ \prod_{\nu=1}^{k} \frac{t_\nu^{2i_\nu}}{(t_\nu - t_{\nu+1})^{3-2i_\nu}} \right\} f(t_1)\, dt_1 \cdots dt_k.$$

Here the i_ν's are either zero or one and all combinations are permitted. Further $m = k - \sum i_\nu$.

As a matter of fact, the k-fold integrals in (1.2) can be replaced by m-fold ones if we use an obvious contraction process based upon Cauchy's integral formula. Each of the integrations with respect to a variable t_ν such that $i_\nu = 1$ can be suppressed if appropriate modifications are made in the remaining integrations. The result is an expression of the form

$$(1.3) \qquad \begin{aligned} J_{j_0 j_1 \cdots j_m} &= (-\pi i)^{-m} z^{2j_0} \\ &\cdot \oint \underset{(m)}{\cdots} \oint \frac{s_1^{2j_1} s_2^{2j_2} \cdots s_m^{2j_m}}{(s_1 - s_2)^3 (s_2 - s_3)^3 \cdots (s_m - z)^3} f(s_1)\, ds_1 \cdots ds_m, \end{aligned}$$

where $j_0, j_1, \cdots, j_m$ are non-negative integers such that

$$(1.4) \qquad j_0 + j_1 + \cdots + j_m = k - m.$$

We have then

$$(1.5) \qquad \delta_z^k f(z) = \sum J_{j_0 j_1 \cdots j_m}.$$

Here the summation extends over all integers m $(0 \leq m \leq k)$ and all integers $j_0, j_1, \cdots, j_m$ subject to (1.4). The contours of integration are subject to the obvious restrictions.

We have shown ([4], p. 892) that the reduction can be carried still further.

For the details we refer to the quoted passage. The result is an expression of the form

$$(1.6)\qquad \begin{aligned} J_{j_0 j_1 \cdots j_m} &= (2\pi i)^{-\mu-1}\left\{\prod_{\alpha=1}^{\mu+1}(d_\alpha)!\right\} z^{2j_0} \\ &\cdot \oint \underset{(\mu+1)}{\cdots} \oint \frac{u_1^{2\nu_1} u_2^{2\nu_2} \cdots u_\mu^{2\nu_\mu} f(u_1)\, du_1 \cdots du_{\mu+1}}{(u_1 - u_2)^{d_1+1}(u_2 - u_3)^{d_2+1} \cdots (u_{\mu+1} - z)^{d_{\mu+1}+1}}. \end{aligned}$$

Here it is supposed that the subscripts $j_1, j_2, \cdots, j_m$ vanish except those in the places $i_1, i_2, \cdots, i_\mu$. Further

$$j_{i_\alpha} = \nu_\alpha, \quad i_0 = 0, \quad i_{\mu+1} = m, \quad 2(i_\alpha - i_{\alpha-1}) = d_\alpha, \quad s_{i_\alpha} = u_\alpha.$$

We note that

$$d_1 + d_2 + \cdots + d_\mu + d_{\mu+1} = 2m.$$

The formula remains valid if all $j_i = 0$. We have then $\mu = 0$, $d_1 = 2m$, $j_0 = k - m$, and a single integral. It is also valid if $i_\mu = m$. We have then $d_{\mu+1} = 0$ and take $(d_{\mu+1})! = 0! = 1$.

For the purpose of making estimates we have found it convenient to resolve $\delta_z^k f(z)$ into its 2^k component integrals. But when it comes to getting representations of differential operators of the type $G(\delta_z)$, it is advantageous to reassemble these components into groups.

Let us return to formula (1.3) and choose contours of integration which may depend upon k and m but not upon the subscripts $j_0, j_1, \cdots, j_m$. Summing all m-fold integrals, we get

$$(-\pi i)^{-m} \oint \cdots \oint \frac{L_{k-m}(z, t_1, t_2, \cdots, t_m)}{(t_1 - t_2)^3 (t_2 - t_3)^3 \cdots (t_m - z)^3} f(t_1)\, dt_1 \cdots dt_m.$$

Here

$$(1.7)\qquad L_\nu(t_0, t_1, \cdots, t_m) = \sum t_0^{2j_0} t_1^{2j_1} \cdots t_m^{2j_m},$$

where the summation extends over all non-negative integers $j_0, j_1, \cdots, j_m$ such that

$$(1.8)\qquad j_0 + j_1 + \cdots + j_m = \nu.$$

We have consequently

$$(1.9)\qquad \delta_z^k f(z) = \sum_{m=0}^{k} (-\pi i)^{-m} \oint \underset{(m)}{\cdots} \oint \frac{L_{k-m}(z, t_1, \cdots, t_m)}{(t_1 - t_2)^3 \cdots (t_m - z)^3} f(t_1)\, dt_1 \cdots dt_m,$$

where the contours of integration can still be disposed within obvious bounds. The first term in the sum corresponding to $m = 0$ is understood to be

$$L_k(z) f(z) \equiv z^{2k} f(z).$$

As an application of formula (1.9), let us compute $\delta_z^k z^n$, where n is a non-negative integer. The integrals are easily worked out and give

$$\delta_z^k z^n = \sum (-1)^m A_{n,k,m} z^{n+2k-4m}, \tag{1.10}$$

where

$$\begin{aligned} A_{n,k,m} = \sum\, &(n + 2j_1)(n + 2j_1 - 1)(n + 2j_1 + 2j_2 - 2) \\ &\cdot(n + 2j_1 + 2j_2 - 3) \cdots (n + 2j_1 + 2j_2 + \cdots + 2j_m - 2m + 2) \\ &\cdot(n + 2j_1 + 2j_2 + \cdots + 2j_m - 2m + 1). \end{aligned} \tag{1.11}$$

Here the summation extends over all non-negative integers $j_1, j_2, \cdots, j_m$ subject to the condition $0 \leqq j_1 + j_2 + \cdots + j_m \leqq k - m$. In (1.10) the summation extends over all non-negative integers m not exceeding the smaller of the numbers k and $\frac{1}{4}(2k + n)$. Evidently $A_{n,k,m}$ is a positive integer, and taking merely the term corresponding to $j_1 = k - m, j_2 = \cdots = j_m = 0$ we get the trivial estimate

$$A_{n,k,m} > \frac{(n + 2k - 2m)!}{(n + 2k - 4m)!} \tag{1.12}$$

which gives some idea of the rate of growth of these coefficients. We see that $\delta_z^k z^n$ is a polynomial of degree $n + 2k$ which reaches its maximum for fixed values of $|z|$ on the lines $y = \pm x, z = x + iy$.

On the basis of these formulas we can get expansions of the δ-transforms of an arbitrary analytic function holomorphic at the origin. Let

$$f(z) = \sum_{n=0}^{\infty} a_n z^n, \quad |z| < R; \qquad G(w) = \sum_{k=0}^{\infty} g_k w^k, \quad |w| < \infty. \tag{1.13}$$

Then

$$\delta_z^k f(z) = \sum_{n=0}^{\infty} a_n \sum_m (-1)^m A_{n,k,m} z^{n+2k-4m}, \tag{1.14}$$

$$G(\delta_z)\cdot f(z) = \sum_{k=0}^{\infty} g_k \sum_{n=0}^{\infty} a_n \sum_m (-1)^m A_{n,k,m} z^{n+2k-4m}. \tag{1.15}$$

Here formula (1.14) is easily justified and the double series is absolutely convergent for $|z| < R$. Formula (1.15), on the other hand, is of more problematic nature. If, however, the operator $G(\delta_z)$ is known to apply to all analytic functions, then (1.15) is a valid representation of the transform for $|z| < R$ if the triple series is summed in the order indicated, that is, first with respect to m, then n and finally k. The same conclusion is of course valid if $f(z)$ belongs to some more restricted class $\mathfrak{F}$ of analytic functions and it is known a priori that the operator $G(\delta_z)$ applies to $\mathfrak{F}$.

Formula (1.15) represents $G(\delta_z)\cdot f(z)$ as a double series in the polynomials $\delta_z^k z^n$. The domain of convergence of a polynomial series of course need not be

a circle, but we have no idea of what domains are possible in the present case. Sometimes it is possible to rearrange the polynomial series into a power series within the circle $|z| < R$. The following is a particularly important case of which we shall make an application below in §8.6.

Suppose that[5]

$$\omega^{2k} g_k \geqq 0, \qquad \omega = e^{\frac{1}{2}\pi i} \quad \text{or} \quad e^{-\frac{1}{2}\pi i}, \tag{1.16}$$

and that $G(\delta_z)$ is known to apply to a class $\mathfrak{F}$ of the following structure. All functions $f(z)$ of $\mathfrak{F}$ are holomorphic at $z = 0$; and if

$$f(z) = \sum_{n=0}^{\infty} a_n z^n$$

belongs to $\mathfrak{F}$, then

$$f^*(z) = \sum_{n=0}^{\infty} a_n^* z^n$$

also belongs to $\mathfrak{F}$ if $|a_n^*| = |a_n|$ for all n. We can then assert that the triple series in (1.15) is absolutely convergent within the circle of holomorphism of $f(z)$ for every $f(z)$ in $\mathfrak{F}$. Indeed, the series converges for every $f(z)$ in $\mathfrak{F}$ by assumption. The series is known to remain convergent if we replace every a_n by $a_n^* = \omega^{-n} |a_n|$. If in this convergent series we put $z = \omega r$ $(0 < r < R)$ and observe (1.16) and the definition of a_n^*, we obtain a triple series all the terms of which are non-negative. This series is simply the series obtained when we replace every term in (1.15) by its absolute value. Consequently the series (1.15) is absolutely convergent within the circle of holomorphism of $f(z)$ for any $f(z)$ in $\mathfrak{F}$. Such a series can of course be rearranged as a power series in z. We do not insist further on the properties of the series (1.15).

Representations of $G(\delta_z)\cdot f(z)$ of more general usefulness can be obtained directly from formula (1.9). We have

$$G(\delta_z)\cdot f(z) = \sum_{k=0}^{\infty} g_k \sum_{m=0}^{k} (-\pi i)^{-m} \oint \cdots \underset{(m)}{} \oint \frac{L_{k-m}(z, t_1, \cdots, t_m)}{(t_1 - t_2)^2 \cdots (t_m - z)^2} f(t_1)\, dt_1 \cdots dt_m. \tag{1.17}$$

Let us suppose that in this formula all m-fold integrals are taken along the same paths of integration, regardless of the value of k, and that the double series is absolutely convergent in such a manner that summation and integration can be interchanged. These hypotheses will be critically examined in §4 below. Introducing the entire functions

$$G_0(z) = \sum_{k=0}^{\infty} g_k L_k(z) \equiv G(z^2), \tag{1.18}$$

$$G_m(z, t_1, \cdots, t_m) = \sum_{k=0}^{\infty} g_k L_{k-m}(z, t_1, \cdots, t_m), \tag{1.19}$$

[5] For the sake of simplicity we have assumed that (1.16) is valid for all k but all large k would be sufficient for our purposes.

we can write

$$
\begin{aligned}
G(\delta_z)\cdot f(z) &= G_0(z)\cdot f(z) \\
&\quad + \sum_{m=1}^{\infty} (-\pi i)^{-m} \oint \cdots \oint_{(m)} \frac{G_m(z, t_1, \cdots, t_m)}{(t_1 - t_2)^3 \cdots (t_m - z)^3} f(t_1)\, dt_1 \cdots dt_m .
\end{aligned} \tag{1.20}
$$

The right side may have a meaning even if the left has none. In this case we use the symbol $G^*(\delta_z)\cdot f(z)$ and regard $G^*(\delta_z)$ as an extension of $G(\delta_z)$. In all cases considered below, however, formula (1.20) will be found to represent $G(\delta_z)\cdot f(z)$ proper.

2. Estimates of $\delta_z^k f(z)$. We shall investigate how the iterates of $\delta_z f(z)$ grow in absolute value with k. Various estimates will be obtained from the different formulas of the preceding section.

We start with formula (1.1). Let us choose as contours of integration the circles

$$| t_\nu - t_{\nu+1} | = \frac{p}{k} \qquad (\nu = 1, 2, \cdots, k;\ t_{k+1} = z),$$

where p is any positive number less than $R(z)$, the radius of holomorphism of $f(z)$ at z. Let $M_z(p)$ denote the maximum modulus of $f(t)$ on the circle $| t - z | = p$ and put $| z | = r$. Then

$$
\begin{aligned}
| \delta_z^k f(z) | &\leqq M_z(p)\left(\frac{k}{p}\right)^{2k} \prod_{\nu=1}^{k} \left\{\left(r + \frac{\nu}{k} p\right)^2 \frac{p^2}{k^2} + 2\right\} \\
&= M_z(p) 2^k \left(\frac{k}{p}\right)^{2k} \prod_{\nu=1}^{k} \left\{1 + \frac{p^2}{2k^2}\left(r + \frac{\nu}{k} p\right)^2\right\} \\
&< M_z(p) 2^k \left(\frac{k}{p}\right)^{2k} \exp\left\{\frac{p^2}{2k^2} \sum_{\nu=1}^{k} \left(r + \frac{\nu}{k} p\right)^2\right\}
\end{aligned}
$$

and finally

$$
\begin{aligned}
| \delta_z^k f(z) | &< M_z(p) 2^k \left(\frac{k}{p}\right)^{2k} \\
&\quad \cdot \exp\left\{\frac{1}{2}\left(\frac{pr}{k}\right)^2 + \frac{1}{2}\frac{p^3 r}{k^2}(k + 1) + \frac{1}{12}\frac{p^4}{k^3}(k + 1)(2k + 1)\right\}.
\end{aligned} \tag{2.1}
$$

This estimate is not particularly good but is nevertheless quite useful. As an example, suppose that $f(z)$ is an entire function such that

$$\lim_{r\to\infty} r^{-2} \log M(r; f) = 0, \qquad M(r; f) = \max_{0 \leq \theta < 2\pi} | f(re^{i\theta}) |. \tag{2.2}$$

Then $M_z(p) \leqq M(r + p; f)$. Let us choose $p = ak^{\frac{1}{2}}$, where a is independent of k and will be disposed of later. Formula (2.2) implies that the k-th root of

$M(r + ak^{\frac{1}{2}}; f)$ tends to unity as $k \to \infty$. It follows that

$$\limsup_{k \to \infty} k^{-1} | \delta_z^k f(z) |^{1/k} \leqq 2a^{-2} \exp\left(\tfrac{1}{6}a^4\right),$$

and this expression reaches its minimum for $a^4 = 3$. Hence

$$\limsup_{k \to \infty} k^{-1} | \delta_z^k f(z) |^{1/k} \leqq 2\left(\frac{e}{3}\right)^{\frac{1}{2}} \tag{2.3}$$

for any entire function satisfying (2.2). This means that the order ρ of the function is at most 2; and if $\rho = 2$ then the function is of minimal type. We are not able to improve on (2.3), but we are fairly sure that the constant on the right is not the best possible.

In passing we notice the estimate

$$| \delta_z^k f(z) | \leqq M_z(p) \left[(r + p)^2 + 2\left(\frac{k}{p}\right)^2 \right]^k \tag{2.4}$$

which is more favorable than (2.1) for small values of k.

Suppose now that $f(z)$ is an entire function of order $\rho > 2$ and type α, i.e.,

$$\limsup_{r \to \infty} r^{-\rho} \log M(r; f) = \alpha. \tag{2.5}$$

We then use formulas (1.3)–(1.6) to get our estimates.[6] Formula (1.5) expresses $\delta_z^k f(z)$ as the sum of 2^k integrals. We separate these into two groups according as $2m \leqq k$ or $> k$. The integrals of the first group we take in the form given by formula (1.3). We let p be a positive quantity, to be disposed of later, and in the integral for $J_{i_0, i_1, \ldots, i_m}$ we use as contours of integration the circles

$$| s_\nu - s_{\nu+1} | = \frac{p}{m} \qquad (\nu = 1, 2, \cdots, m; s_{m+1} = z).$$

We note that $| s_\nu | \leqq r + p$ for all ν. Hence

$$| J_{i_0, i_1, \ldots, i_m} | \leqq M(r + p; f) 2^m (r + p)^{2(k-m)} \left(\frac{m}{p}\right)^{2m}. \tag{2.6}$$

This estimate is also true for $m = 0$ if we replace the meaningless 0^0 by 1. Let the sum of the terms $J_{i_0, i_1, \ldots, i_m}$ of the first group in which $2m \leqq k$ be denoted by S_k^1. We have then

$$| S_k^1 | \leqq M(r + p; f)(r + p)^{2k} \sum_{2m \leqq k} \binom{k}{m} 2^m \left[\frac{m}{p(r + p)} \right]^{2m},$$

[6] For the subsequent discussion cf. E. Hille [4], pp. 892–894. The estimate of S_k^1 is new and the factor $2e^{\frac{1}{2}}$ in the exponent in (2.10) is better than the factor 4 previously found ([4], formula (3.3.11)), but it is still not the best possible factor. This affects formula (2.16) below adversely.

since there are $\binom{k}{m}$ m-fold integrals. The function $(x/A)^x$ is less than or equal to 1 for $0 < x \leqq A$. Further

$$\sum_{2m\leqq k} \binom{k}{m} 2^m < 2^{\frac{1}{2}k} \sum_0^k \binom{k}{m} = 2^{\frac{3}{2}k}.$$

Hence

$$|S_k^1| \leqq M(r+p;f)[2^{\frac{3}{4}}(r+p)]^{2k} \max\left\{1, \left[\frac{k}{2p(r+p)}\right]^k\right\}. \tag{2.7}$$

Let us denote the sum of the integrals $J_{j_0,j_1,\cdots,j_m}$ with $2m > k$ by S_k^2. Here we use the reduced form (1.6). Let q be another positive quantity to be disposed of later and choose as contours of integration the circles

$$|u_\alpha - u_{\alpha+1}| = \frac{d_\alpha q}{2m} \qquad (\alpha = 1, 2, \cdots, \mu+1;\ u_{\mu+2} = z).$$

By the usual methods we get

$$|J_{j_0,j_1,\cdots,j_m}| \leqq M(r+q;f)(r+q)^{2(k-m)} \left(\frac{2m}{q}\right)^{2m} \prod_\alpha \left\{\frac{(d_\alpha)!}{d_\alpha^{d_\alpha}}\right\}. \tag{2.8}$$

It was shown ([4], pp. 893–894) that

$$|S_k^2| \leqq M(r+q;f) 4 \left(\frac{2k}{eq}\right)^{2k} \left\{k^{\frac{1}{2}} + \sum_{j=1}^{\frac{1}{2}k} \frac{[eq(r+q)]^{2j}}{j^j j!}\right\}. \tag{2.9}$$

The expression within the braces is dominated by a suitably chosen exponential function, and a simple calculation shows that

$$|S_k^2| \leqq 8M(r+q;f)k^{\frac{1}{2}} \left(\frac{2k}{eq}\right)^{2k} \exp\{2e^{\frac{1}{2}}q(r+q)\}. \tag{2.10}$$

Combining (2.7) and (2.10) we get

$$\begin{aligned} |\delta_z^k f(z)| \leqq{}& M(r+p;f)[2^{\frac{3}{4}}(r+p)]^{2k} \max\left\{1, \left[\frac{k}{2p(r+p)}\right]^k\right\} \\ &+ 8M(r+q;f)k^{\frac{1}{2}} \left(\frac{2k}{eq}\right)^{2k} \exp\{2e^{\frac{1}{2}}q(r+q)\}, \end{aligned} \tag{2.11}$$

where p and q are arbitrary positive numbers.

It is clear that the first term in (2.11) cannot be made essentially less than k^k for large values of k, $|z|$ being fixed, no matter how p is chosen. The second term is much more affected by the choice of q.

So far we have made no use of hypothesis (2.5) so that formula (2.11) is valid for any entire function without restrictions on p and q. Let us now use (2.5).

We choose $p = 1$, $q = bk^{1/\rho}$, where b will be disposed of later. For $k \geqq 2(r + 1)$ the first term in (2.11) becomes

$$M(r + 1; f)[2^{\frac{1}{2}}(r + 1)k]^k. \tag{2.12}$$

Let us now define σ by the order relation

$$\frac{1}{\rho} + \frac{1}{2\sigma} = 1. \tag{2.13}$$

Then the second term in (2.11) may be written

$$8M(r + bk^{1/\rho}; f)k^{\frac{1}{2}+k/\sigma}\left(\frac{2}{be}\right)^{2k} \exp\,[2e^{\frac{1}{2}}bk^{1/\rho}(r + bk^{1/\rho})]. \tag{2.14}$$

Since $\sigma < 1$, the expression (2.14) evidently completely dominates (2.12). By (2.5)

$$\limsup_{k\to\infty}\,[M(r + bk^{1/\rho}; f)]^{1/k} \leqq \exp\,[\alpha b^\rho].$$

Consequently

$$\limsup_{k\to\infty}\,k^{-1/\sigma}\,|\,\delta_z^k f(z)\,|^{1/k} \leqq \left(\frac{2}{be}\right)^2 \exp\,[\alpha b^\rho]$$

for every positive b. Minimizing the right side, we get

$$\limsup_{k\to\infty}\,k^{-1/\sigma}\,|\,\delta_z^k f(z)\,|^{1/k} \leqq \left(\frac{2}{e}\right)^{1/\sigma}(\alpha\rho)^{2/\rho}. \tag{2.15}$$

We shall see in §8.1 that this estimate is the best of its kind in a certain sense.

The case $\rho = 2$, $\alpha > 0$, remains. Here the two terms of (2.11) become essentially of the same order of magnitude and the exponential factor in the second term also affects the estimate. We now choose $p = ak^{\frac{1}{2}}$, $q = bk^{\frac{1}{2}}$, where a and b are to be disposed of later. Assuming $2a^2 < 1$, we have

$$\begin{aligned}|\,\delta_z^k f(z)\,| \leqq{}& M(r + ak^{\frac{1}{2}}; f)2^{\frac{1}{2}k}\left\{1 + \frac{r}{ak^{\frac{1}{2}}}\right\}^k k^k \\ &+ 8M(r + bk^{\frac{1}{2}}; f)\left(\frac{2}{be}\right)^{2k} \exp\,\{2e^{\frac{1}{2}}bk^{\frac{1}{2}}(r + bk^{\frac{1}{2}})\}\,k^{k+\frac{1}{2}}.\end{aligned}$$

Since $(A + B)^\gamma < A^\gamma + B^\gamma$ when $\gamma < 1$, we get

$$\begin{aligned}k^{-1}\,|\,\delta_z^k f(z)\,|^{1/k} \leqq{}& [M(r + ak^{\frac{1}{2}}; f)]^{1/k}2^{\frac{1}{2}}\left\{1 + \frac{r}{ak^{\frac{1}{2}}}\right\} \\ &+ [8M(r + bk^{\frac{1}{2}}; f)]^{1/k}\left(\frac{2}{be}\right)^2 \exp\left\{2e^{\frac{1}{2}}b^2\left(1 + \frac{r}{bk^{\frac{1}{2}}}\right)\right\}k^{1/(2k)} \\ &\to 2^{\frac{1}{2}} \exp\,(\alpha a^2) + \left(\frac{2}{be}\right)^2 \exp\,[b^2(\alpha + 2e^{\frac{1}{2}})]\end{aligned}$$

as $k \to \infty$. This is true for all $a > 0$ so we can let $a \to 0$, and it is also true for all values of $b > 0$ so we can choose $b^2 = (\alpha + 2e^{\frac{1}{2}})^{-1}$ which minimizes. Hence we get

$$\limsup_{k\to\infty} k^{-1} | \delta_z^k f(z) |^{1/k} \leqq \frac{4}{e}\alpha + \frac{8}{e^{\frac{1}{2}}} + 2^{\frac{1}{2}}. \tag{2.16}$$

No claim is made that this estimate is the best possible, but we shall prove in §8.1 that the factor $4/e$ in front of α cannot be replaced by any smaller quantity.

For small values of α formula (2.16) gives far too high an estimate. For such values we get a much better result by following the method used in deriving formula (2.3). This method gives

$$\limsup_{k\to\infty} k^{-1} | \delta_z^k f(z) |^{1/k} \leqq 2a^{-2} \exp [\alpha a^2 + \tfrac{1}{6}a^4] \tag{2.17}$$

for all real values of a. The minimum of the right side is reached for

$$a^2 = -\tfrac{3}{2}\alpha + (\tfrac{9}{4}\alpha^2 + 3)^{\frac{1}{2}}. \tag{2.18}$$

For small values of α this estimate gives a limit of the form

$$2\left(\frac{e}{3}\right)^{\frac{1}{2}} \{1 + 3^{\frac{1}{2}}\alpha + O(\alpha^2)\}. \tag{2.19}$$

For large values of α it gives

$$2e\alpha + O\left(\frac{1}{\alpha^2}\right). \tag{2.20}$$

Here (2.19) is much better than (2.16), while (2.20) is not so good as (2.16). We do not insist on further refinements of these estimates. The results obtained so far can be summarized as follows. We recall that statements regarding "best possible" estimates will be proved in §8.1.

THEOREM 2.1. *If $f(z)$ is an entire function of order ρ and type α, then*

(i) $$\limsup_{k\to\infty} k^{-1/\sigma} | \delta_z^k f(z) |^{1/k} \leqq \left(\frac{2}{e}\right)^{1/\sigma} (\alpha\rho)^{2/\rho}$$

when $\rho > 2$, where the conjugate exponent σ is determined by the order relation (2.13). This estimate is the best of its kind.

(ii) $$\limsup_{k\to\infty} k^{-1} | \delta_z^k f(z) |^{1/k} \leqq \min \left\{\frac{4}{e}\alpha + \frac{8}{e^{\frac{1}{2}}} + 2^{\frac{1}{2}}, 2a^{-2} \exp [\alpha a^2 + \tfrac{1}{6}a^4]\right\}$$

when $\rho = 2$, where a^2 is defined by (2.18). For large values of α the factor $4/e$ cannot be replaced by any smaller number.

(iii) $$\limsup_{k\to\infty} k^{-1} | \delta_z^k f(z) |^{1/k} \leqq 2\left(\frac{e}{3}\right)^{\frac{1}{2}}$$

when $\rho < 2$ or $\rho = 2$ and $\alpha = 0$.

In the discussion above we have restricted ourselves to entire functions of normal type or minimal type. More precise information could be obtained by the introduction of the so-called *proximate orders* of Boutroux and Lindelöf in both these cases and also in the case of functions of maximal type.[7] We leave such extensions to the interested reader.

3. **On a class of functionals.** The results obtained in the preceding section can be formulated in a slightly different manner which is of some interest. Let $f(z)$ be an analytic function holomorphic in some domain D. Let z be a fixed point of D, λ a real number $0 < \lambda \leqq 1$, and form

$$F_\lambda(z;f) \equiv \limsup_{k\to\infty} \left[\frac{|\delta_z^k f(z)|}{\Gamma\left(1+\frac{k}{\lambda}\right)}\right]^{1/(2k)}, \tag{3.1}$$

where we admit $+\infty$ as a possible value. This defines a non-negative function of z in D. *As a function of z it is not necessarily continuous in D even if it is bounded.* An example showing this will be found in §8.2.

But we can also regard $F_\lambda(z;f)$ for fixed z in D as a *functional* of the second argument defined for all functions holomorphic at the point in question. As such it is evidently *non-linear*, but it is *quasi-additive*

$$F_\lambda(z;f_1+f_2) \leqq F_\lambda(z;f_1) + F_\lambda(z;f_2), \tag{3.2}$$

and for every constant $C \neq 0$

$$F_\lambda(z;Cf) = F_\lambda(z;f). \tag{3.3}$$

These functionals are closely related to various types of *grades*[8] which have been considered in the theory of linear operations.

Now let $\mathfrak{F}_{\rho,\alpha}$ denote the class of all entire functions $f(z)$ such that

$$\limsup_{r\to\infty} r^{-\rho} \log M(r;f) \leqq \alpha. \tag{3.4}$$

We can then reformulate Theorem 2.1 as follows:

THEOREM 3.1. *The functional $F_\lambda(z;f)$ has the following properties when $f(z)$ ranges over $\mathfrak{F}_{\rho,\alpha}$. $F_\lambda(z;f) \equiv 0$ for $\lambda < \sigma$ when $\rho > 2$, and for $\lambda < 1$ when $\rho \leqq 2$. For $\rho > 2$*

[7] See G. Valiron [9], §III, 6.

[8] This is the term used by H. T. Davis ([2], Chapter V, *Grades defined by special operators*) for the superior limit of the k-th root of the absolute value of the transform of the k-th power of the operator. This corresponds to the case $\sigma = \infty$ in (3.1) which seems to be of limited interest to us. I. M. Sheffer used the term exponential value; the German term "Stufe" was introduced by O. Perron. The notion itself in one form or another goes back to C. Bourlet and S. Pincherle. If the grade is infinite, H. T. Davis discusses how fast the sequence in question tends to infinity under different assumptions on the function. The introduction of the Gamma function in the functional puts the investigation on a systematic basis in our case. More general grades have been considered by P. Flamant [3].

$$F_\sigma(z;f) \leqq (\alpha\rho)^{1/\rho}(2\sigma)^{1/(2\sigma)}, \tag{3.5}$$

where σ is the conjugate order of ρ, and no better estimate is valid for the class $\mathfrak{F}_{\rho,\alpha}$. $F_1(z;f)$ is uniformly bounded on $\mathfrak{F}_{\rho,\alpha}$ when $\rho \leqq 2$, and bounds can be read off from (ii) *and* (iii) *of Theorem* 2.1. *Finally, $F_\lambda(z;f)$ is unbounded on $\mathfrak{F}_{\rho,\alpha}$ for $\lambda > \sigma$ when $\rho > 2$, and for $\lambda > 1$ when $\rho \leqq 2$.*

Let us now investigate the continuity properties of $F_\lambda(z;f)$ as a functional of $f(z)$ on $\mathfrak{F}_{\rho,\alpha}$. We shall say that $f_n(z)$ converges to $f(z)$ in $\mathfrak{F}_{\rho,\alpha}$ if $f_n(z) \in \mathfrak{F}_{\rho,\alpha}$ for all n, $f(z) \in \mathfrak{F}_{\rho,\alpha}$, and $f_n(z)$ converges uniformly to $f(z)$ in every fixed circle $|z| \leqq R$ of the z-plane. We then say as usual that $F_\lambda(z;f)$ is continuous at $f = f_0$, if $f_n \to f_0$ implies $F_\lambda(z;f_n) \to F_\lambda(z;f_0)$. Since $F_\lambda(z;f)$ is a non-negative quasi-additive functional which vanishes for $f = 0$, continuity anywhere implies continuity everywhere. Starting with this remark, we shall prove

THEOREM 3.2. *$F_\lambda(z;f)$ is not continuous anywhere on $\mathfrak{F}_{\rho,\alpha}$ for $\lambda = \sigma$ when $\rho > 2$ and for $\lambda = 1$ when $\rho \leqq 2$.*

For the proof it is enough to exhibit a sequence f_n converging to zero in $\mathfrak{F}_{\rho,\alpha}$, such that $F_\lambda(z;f_n)$ does not converge to zero at least for some values of z, where λ is σ or 1. The existence of such functions is an immediate consequence of formula (3.3). We have merely to choose a function $f(z)$ such that $F_\lambda(z;f) \not\equiv 0$ for $\lambda = \sigma$ or 1 respectively and then take $f_n(z) = C_nf(z)$, where $C_n \to 0$. The possibility of finding such a function $f(z)$ will be established in §8.1.

We have proved elsewhere and it will be proved again in §4 that $F_{\frac{1}{2}}(z;f)$ is definable over the class of all analytic functions. More precisely, it was shown that

$$F_{\frac{1}{2}}(z;f) \equiv \limsup_{k\to\infty} \left[\frac{|\delta_z^k f(z)|}{(2k)!}\right]^{1/(2k)} \leqq \frac{1}{R(z)}, \tag{3.6}$$

where $R(z)$ is the distance from z to the nearest singular point of $f(z)$, and that moreover this inequality is the best of its kind.[9] Judging from the analogy with the functional

$$\Phi_1(z;f) \equiv \limsup_{k\to\infty} \left[\frac{|f^{(k)}(z)|}{k!}\right]^{1/k} = \frac{1}{R(z)}, \tag{3.7}$$

one would imagine that the sign of equality should always hold in (3.6) for every f and every point z. Unfortunately this is not the case and it is even obvious that it could not be so.[10] Indeed, if $f^{(2n)}(0) = 0$ for all n, then $F_{\frac{1}{2}}(0;f) = 0$ regardless of the value of $R(0)$. Moreover, $F_{\frac{1}{2}}(z;f)$ *can be a discontinuous function of z inside the domain of holomorphism of $f(z)$ and it is an everywhere discontinuous functional over the class of analytic functions $f(z)$.* An example to prove the first point will be found in §8.2. The second statement is proved as Theorem 3.2.

[9] See [4], formula (3.3.13) and Theorem 3.1. Our present formula (3.6) gives a more pregnant formulation of the result. That the inequality is the best possible also follows from the example in §8.2, especially formula (8.2.3).

[10] This does not exclude the possibility that equality may hold almost everywhere, for instance, or that the origin is the only exceptional point.

4. Entire functions of δ_z. Let

$$G(w) = \sum_{k=0}^{\infty} g_k w^k$$

be an entire function of w, the order of which will be restricted below. We proceed now to a systematic study of the differential operator $G(\delta_z)$. We recall the definition of applicability given in the introduction: *$G(\delta_z)$ applies to the class $\mathfrak{F}$ of analytic functions $f(z)$ if the series*

$$G(\delta_z)\cdot f(z) = \sum_{k=0}^{\infty} g_k \delta_z^k f(z) \tag{4.1}$$

converges to a holomorphic function at every point where $f(z)$ is holomorphic, regardless of what function $f(z) \in \mathfrak{F}$ we take.

In §1 formula (1.20) was given as a formal representation of the transform

$$\begin{aligned} G(\delta_z)\cdot f(z) &= G_0(z)\cdot f(z) \\ &+ \sum_{m=1}^{\infty} (-\pi i)^{-m} \oint \cdots \oint \frac{G_m(z, t_1, t_2, \cdots, t_m)}{(t_1 - t_2)^3 \cdots (t_m - z)^3} f(t_1)\, dt_1 \cdots dt_m, \end{aligned} \tag{4.2}$$

where the contours of integration in the m-th term are circles $| t_\nu - t_{\nu+1} | = R_{m,\nu}$ such that $\sum_{\nu=1}^{m} R_{m,\nu} < R(z)$, the radius of holomorphism of $f(z)$ at z. The entire functions $G_m(z, t_1, \cdots, t_m)$ are defined by (1.7), (1.18), and (1.19). Now formula (4.2) is obtained by substituting the expression for $\delta_z^k f(z)$ furnished by formula (1.9) into (4.1) and rearranging terms. This is evidently permitted if (4.2) remains convergent when every quantity is replaced by its absolute value and, in addition, $| G_m(z, t_1, \cdots, t_m) |$ by $\sum | g_k || L_{k-m}(z, t_1, \cdots, t_m) |$. All our convergence proofs will involve such replacements so we can rest assured that if the resulting series are convergent, they do give valid representations of the transform and the operator $G(\delta_z)$ does apply to the class of functions under consideration.

We shall begin the discussion by an investigation of the entire functions $G_m(z, t_1, \cdots, t_m)$. We shall suppose that

$$\limsup_{r\to\infty} r^{-\sigma} \log \max_{0\leq\theta<2\pi} | G(re^{i\theta}) | \leq \beta, \tag{4.3}$$

restricting ourselves for reasons which will become apparent below to the case in which $0 < \sigma \leq 1$.[11] Let $\mathfrak{G}_{\sigma,\beta}$ denote the class of all such functions $G(w)$.[12]

[11] The case $\sigma = 0$ would require the introduction of proximate orders or some such device and is excluded for convenience. Only the range $\frac{1}{2} \leq \sigma \leq 1$ is of interest below.

[12] The reader should observe that the classes $\mathfrak{F}_{\rho,\alpha}$ and $\mathfrak{G}_{\sigma,\beta}$ become identical if $\rho = \sigma$, $\alpha = \beta$, $z = w$. It is convenient, however, to distinguish between a *space of operands* and a *space of operators* and our notation is chosen accordingly. Thus the letters f, z, $\mathfrak{F}$, ρ, α always refer to the operand space and G, w, $\mathfrak{G}$, σ, β always to the operator space.

LEMMA 4.1. *If $G(w) \in \mathfrak{G}_{\sigma,\beta}$ and $|z|, |t_1|, \cdots, |t_m| \leqq u$, then for every $\epsilon > 0$*

$$|G_m(z, t_1, \cdots, t_m)| \leqq C[\sigma e(\beta + \epsilon)/m]^{m/\sigma} \exp[(\beta + \epsilon)u^{2\sigma}], \tag{4.4}$$

where C depends upon β, ϵ, and σ, but not upon m or u.

Proof. By assumption[13]

$$|g_k| \leqq C(\epsilon) \frac{(\beta + \epsilon)^{k/\sigma}}{\Gamma(1 + k/\sigma)}.$$

Since

$$|L_{k-m}(z, t_1, \cdots, t_m)| \leqq \binom{k}{m} u^{2(k-m)},$$

we have

$$\begin{aligned} |G_m(z, t_1, \cdots, t_m)| &\leqq C(\epsilon) \sum_{k=m}^{\infty} \binom{k}{m} \frac{(\beta + \epsilon)^{k/\sigma}}{\Gamma(1 + k/\sigma)} u^{2(k-m)} \\ &= \frac{1}{m!} C(\epsilon)(\beta + \epsilon)^{m/\sigma} E_{1/\sigma}^{(m)}((\beta + \epsilon)^{1/\sigma} u^2), \end{aligned} \tag{4.5}$$

where

$$E_\alpha(z) = \sum_{k=0}^{\infty} \frac{z^k}{\Gamma(1 + \alpha k)}$$

is the Mittag-Leffler E_α-function. For large positive x it is known that

$$E_{1/\sigma}(x) = \sigma \exp(x^\sigma) + O\left(\frac{1}{x}\right). \tag{4.6}$$

We can consequently find a quantity $B(\sigma)$ which is bounded for $0 < \epsilon \leqq \sigma \leqq 1$, such that for all $x \geqq 0$

$$E_{1/\sigma}(x) \leqq B(\sigma) \exp(x^\sigma).$$

By Cauchy's formula

$$\frac{1}{m!} E_{1/\sigma}^{(m)}(x) = \frac{1}{2\pi i} \int_{|z-x|=R} \frac{E_{1/\sigma}(z)\, dz}{(z - x)^{m+1}},$$

the right side of which does not exceed

$$R^{-m} E_{1/\sigma}(x + R) \leqq B(\sigma) R^{-m} \exp[(x + R)^\sigma] < B(\sigma) \exp(x^\sigma) R^{-m} \exp(R^\sigma)$$

since $\sigma \leqq 1$. The least value of the last expression is obtained for $R^\sigma = m/\sigma$.

[13] See, for instance, G. Valiron [9], p. 41. The constant $C(\epsilon)$ depends upon $G(w)$ and the choice of ϵ. Similarly, with other constants below. If it is of some importance that the constants are independent of the functions, attention will be called to the fact.

Hence

$$\frac{1}{m!}E_{1/\sigma}^{(m)}(x) < B(\sigma)\left(\frac{\sigma e}{m}\right)^{m/\sigma} \exp(x^\sigma).$$

If we combine this estimate with (4.5), formula (4.4) results and Lemma 4.1 is proved.

Formula (4.4) is the basic inequality in most of the subsequent discussion. It does not lead to quite as sharp estimates as those based on §2. This is in the main because of our replacing all variables $z, t_1, \cdots, t_m$ by the absolute value of the largest among them. In our applications it is known that $|t_j| \leqq |z| + (m-j)q(m)$, where $q(m) \to 0$ and $mq(m) \to \infty$ as $m \to \infty$. We consequently get $u = |z| + mq(m)$ which, when substituted into (4.4), leads to an unnecessarily high estimate. The excess becomes appreciable only when the orders of $f(z)$ and $G(w)$ are conjugate in the sense defined in the introduction. Since this case can be adequately handled by the methods of §2, we shall not attempt to improve upon (4.4).

We are now ready to start with applicability questions. When $\mathfrak{F}$ is the class of all analytic functions, the following theorem has already been stated and proved by the author.[14]

Theorem 4.1. *A necessary and sufficient condition that $G(\delta_z)$ shall be applicable to the class of all analytic functions is that $G(w) \in \mathfrak{G}_{\frac{1}{2},0}$.*

It is easy to give a proof of the sufficiency of this condition on the basis of Lemma 4.1. Let $f(z)$ be an analytic function holomorphic at the point z_0, where the radius of holomorphism is to be $R(z_0)$. In the m-fold integral of formula (4.2) we choose as contours of integration the circles $|t_\nu - t_{\nu+1}| = p/m$, where p is any number less than $R(z_0)$. We take $z \doteq z_0$, put $r = |z_0|$, and choose $u = r + p$, $\beta = 0$, and $\sigma = \frac{1}{2}$ in Lemma 4.1. We then obtain the following majorant of the series (4.2):

$$(4.7) \qquad |G(z_0^2)||f(z_0)| + C(\epsilon)M(p)\exp[\epsilon(r+p)]\sum_{m=1}^{\infty}\frac{1}{2^m}\left(\frac{\epsilon e}{p}\right)^{2m},$$

where $M(p)$ is the maximum modulus of $f(z)$ on the circle $|z - z_0| = p$. Since ϵ is at our disposal and can be chosen less than $2^{\frac{1}{2}}p/e$ a priori, we see that the majorant series is convergent. Moreover, the convergence is evidently uniform if z_0 is restricted to a bounded domain in which $f(z)$ is holomorphic and $R(z_0)$ has a positive infimum. Hence the transform is holomorphic at every finite point where $f(z)$ is holomorphic. This completes the proof of the sufficiency of the condition. For the necessity we refer to the passage quoted in footnote 14. The argument just given evidently also proves the following result.[15]

[14] [4], Theorem 3.2. Our previous proof was based upon a formula of type (2.11).

[15] This theorem can be generalized. Assuming the distance of D_2 to the complement of D_1 to be $\beta + \eta$, we can allow $G(w) \in \mathfrak{G}_{\frac{1}{2},\beta}$, the conclusion being unchanged. The proof

THEOREM 4.2. *Let $D_1 \supset D_2$ be bounded domains in the complex plane such that D_2 has a positive distance η from the complement of D_1. Let $\mathfrak{F}(D_1) \subset \mathfrak{F}(D_2)$ be the classes of all functions holomorphic and bounded in D_1 and D_2 respectively. Let $G(w) \in \mathfrak{G}_{1,0}$. Then $G(\delta_z)$ defines a linear, bounded and consequently continuous transformation on $\mathfrak{F}(D_1)$ to $\mathfrak{F}(D_2)$.*

Indeed, choosing $p = \eta$ and $\epsilon = \eta/e$ in (4.7) we get

$$\max_{z \in D_2} |G(\delta_z) \cdot f(z)| \leq \max_{z \in D_2} \{|G(z^2)| + C_1(\eta) \exp [\eta/e|z|]\} \cdot \max_{z \in D_1} |f(z)|. \tag{4.8}$$

This inequality shows that we are dealing with a bounded transformation of one normed linear vector space upon another.

$G(\delta_z) \cdot f(z)$ is a bilinear transformation involving two function spaces $\mathfrak{F}$ and $\mathfrak{G}$. That it is continuous on $\mathfrak{F}$ when $G(w)$ is fixed in $\mathfrak{G}_{1,0}$ is expressed by Theorem 4.2. We have also continuity on $\mathfrak{G}_{1,0}$ for fixed $f(z)$ in $\mathfrak{F}$. This does not follow directly from (4.8), and we shall prove continuity only for a particular type of convergence which will be referred to as *dominated convergence in* $\mathfrak{G}_{1,0}$. We shall prove

THEOREM 4.3. *Let $\{\Gamma_n(w)\}$ be a sequence of functions in $\mathfrak{G}_{1,0}$ which converges to a function $G(w)$ and is such that there exists a fixed function $\Gamma_0(w)$ in $\mathfrak{G}_{1,0}$ with $M(r; \Gamma_n) \leq M(r; \Gamma_0)$ for all n and r. Let $f(z)$ be holomorphic in a domain D. Then $\Gamma_n(\delta_z) \cdot f(z)$ converges to $G(\delta_z) \cdot f(z)$ in D and the convergence is uniform in any domain $D_0 \subset D$ which is bounded and has a positive distance from the complement of D.*

The dominated convergence of $\Gamma_n(w)$ to $G(w)$ evidently implies uniform convergence in any finite domain and also implies that $G(w) \in \mathfrak{G}_{1,0}$.[16] Hence we have also

$$\lim_{n \to \infty} \Gamma_{n,m}(z, t_1, \cdots, t_m) = G_m(z, t_1, \cdots, t_m)$$

for every fixed m, and the convergence is uniform when the variables are bounded. Further, the existence of a common dominant of the sequence of maximal moduli implies that the estimates of Lemma 4.1 hold uniformly with respect to n for

cannot be based upon formula (4.2) and Lemma 4.1, which would give a weaker result, but requires the more powerful machinery of §2. Compare G. Pólya [6], p. 600, for the operator $G(d/dz)$. Pólya assumes that D_2 has no points in common with the point-set obtained by adding the conjugate indicator diagram of $G(w)$, in his case a function of exponential type, to the complement of D_1. I do not know if the indicator diagram can be worked into the theory of the operator $G(\delta_z)$.

[16] The first statement follows from Vitali's theorem.—Uniform convergence relative to the function $M(r; \Gamma_0)$ in the sense of E. H. Moore implies but is not implied by dominated convergence. Relatively uniform convergence with respect to an "étalonnage" has been used by P. Flamant [3] in operator theory, but dominated convergence is possibly new in this connection. Cf., however, J. F. Ritt [7], p. 30 et seq.—Theorem 4.3 extends to operators in $\mathfrak{G}_{\sigma,\beta}$ ($\frac{1}{2} < \sigma \leq 1$) and entire functions $f(z)$.

all the functions $\Gamma_{n,m}(z, t_1, \cdots, t_m)$. It follows that for $z \in D_0$, $p = 2^{-\frac{1}{2}}\eta$, $\epsilon = \frac{1}{2}\eta e^{-1}$ we have

$$\Big| \Gamma_n(\delta_z)\cdot f(z) - \Gamma_n(z^2)\cdot f(z) - \sum_{m=1}^{N} (-\pi i)^{-m} \cdot \oint \cdots \oint \frac{\Gamma_{n,m}(z, t_1, \cdots, t_m)}{(t_1 - t_2)^3 \cdots (t_m - z)^3} f(t_1)\, dt_1 \cdots dt_m \Big|$$

$$\leqq 4^{-N} C_2(\eta) \exp\left(\frac{\eta}{2e}|z|\right) \max_{z \in D_1} |f(z)|,$$

where D_1 is the subset of D the points of which have a distance $\leqq 2^{-\frac{1}{2}}\eta$ from some point of D_0. This implies that

$$\limsup_{n\to\infty} |\Gamma_n(\delta_z)\cdot f(z) - G(\delta_z)\cdot f(z)| \leqq 2\cdot 4^{-N} C_2(\eta) \exp(\tfrac{1}{2}\eta e^{-1}|z|) \max_{z \in D_1} |f(z)|$$

for every N, whence the theorem follows.

THEOREM 4.4. *A necessary and sufficient condition that $G(\delta_z)$ shall apply to the class of all entire functions is that $G(w) \in \mathfrak{G}_{\frac{1}{2},\beta}$ for some finite β.*[17]

The necessity will follow from the example in §8.3. For the sufficiency proof, assume that (4.3) holds with $\sigma = \frac{1}{2}$ and a finite β. In formula (4.2) we choose as contours of integration in the m-fold integral the circles $|t_\nu - t_{\nu+1}| = q(m)$, where $q(m)$ is to be chosen suitably. Taking $u = r + mq(m)$, $\sigma = \frac{1}{2}$ in Lemma 4.1 we obtain the following dominant of the series (4.2)

$$|G(z^2)||f(z)| + C(\epsilon) \sum_{m=1}^{\infty} \left[\frac{e(\beta + \epsilon)}{2^{\frac{1}{2}} mq(m)}\right]^{2m} \exp[(\beta + \epsilon)(r + mq(m))] M(r + mq(m); f),$$

where $r = |z|$ and $M(r; f) = \max |f(z)|$ when $|z| = r$.

No matter how fast $M(r; f)$ tends to infinity with r, we can choose $q(m)$ subject to the following conditions:

(1) $q(m) \to 0$ as $m \to \infty$,
(2) $mq(m) \to \infty$,
(3) $\limsup_{m\to\infty} [M(r + mq(m); f)]^{1/m} \leqq 2$.

We see then that the series (4.8) is uniformly convergent in any fixed circle $|z| \leqq R$, since the m-th root of the m-th term tends to zero. It follows that $G(\delta_z)\cdot f(z)$ exists and is an entire function. This completes the proof of the sufficiency of the condition of the theorem.

5. The order relation, conjugate orders and types. The classes $\mathfrak{F}_{\rho,\alpha}$ and $\mathfrak{G}_{\sigma,\beta}$ are characterized by the inequalities

$$\limsup_{r\to\infty} r^{-\rho} \log \max_{0 \leqq \theta < 2\pi} |f(re^{i\theta})| \leqq \alpha, \tag{5.1}$$

[17] This theorem was announced in [4], footnote 10, p. 899.

$$(5.2) \qquad \limsup_{r\to\infty} r^{-\sigma} \log \max_{0\le\theta<2\pi} |G(re^{i\theta})| \leqq \beta.$$

We say that ρ and σ are *conjugate orders* if

$$(5.3) \qquad \sigma = 1, \qquad \rho < 2,$$

$$(5.4) \qquad \frac{1}{\rho} + \frac{1}{2\sigma} = 1, \qquad \rho \geqq 2.$$

We call (5.4) the *order relation.*[18] The value $\rho = 2$ below which the order relation ceases to hold is the *critical ρ-order*, and the corresponding value $\sigma = 1$ is the *maximal σ-order*.

Suppose that $\rho > 2$ and σ is the conjugate order of ρ. We then say that α and β are *conjugate types* if

$$(5.5) \qquad (\alpha\rho)^{1/\rho}(2\beta\sigma)^{1/(2\sigma)} = 1.$$

The importance of these notions will become evident in the following. We shall begin by proving

THEOREM 5.1. *A necessary and sufficient condition that $G(\delta_z)$ shall be applicable to all classes $\mathfrak{F}_{\rho,\alpha}$ with ρ fixed $\geqq 2$ $(0 \leqq \alpha < \infty)$ is that $G(w) \in \mathfrak{G}_{\sigma,0}$, where σ is the conjugate order of ρ. If $\rho < 2$, the condition is merely sufficient.*[19]

At this point we shall prove only the sufficiency of the condition. The necessity will follow from the examples in §8.4. Let us assume then that $\rho \geqq 2$, σ is the conjugate of ρ, and $G(w) \in \mathfrak{G}_{\sigma,0}$. We use formula (1.20) = (4.2), choosing as contours of integration in the m-fold integral the circles

$$(5.6) \qquad |t_\nu - t_{\nu+1}| = m^{-1/(2\sigma)} \qquad (\nu = 1, 2, \cdots, m;\ t_{m+1} = z).$$

We can then substitute $\beta = 0$,

$$(5.7) \qquad u = r + m^{1/\rho}, \qquad |z| = r,$$

in Lemma 4.1. It follows that the series for $G(\delta_z)\cdot f(z)$ is dominated by the series

$$|G(z^2)||f(z)| + C(\epsilon,\sigma) \sum_{m=1}^{\infty} 2^m \left(\frac{\epsilon\sigma e}{m}\right)^{m/\sigma} \cdot m^{m/\sigma} \cdot \exp\,[\epsilon(r + m^{1/\rho})^{2\sigma}] M(r + m^{1/\rho}; f).$$

By assumption (5.1) holds with a finite α. We can then find finite quantities A and B such that

$$M(r;f) < B \exp\,(Ar^\rho)$$

[18] The terminology is mine. Cf. H. Muggli [5], p. 152, for the corresponding relation in the case of the operator $G(d/dz)$. The basic facts appear to be due to G. Valiron [10], pp. 52–53. See also formula (9.3) below.

[19] If we demand instead that $G(\delta_z)$ shall apply to all entire functions of order $\leqq \rho$, including functions of maximal type of order ρ, then it is necessary and sufficient that the order of $G(w)$ be $< \sigma$. See H. Muggli, [5], p. 152, for $G(d/dz)$.

for $r \geqq 1$. The series is consequently dominated by

$$(5.8)\quad |G(z^2)|\,|f(z)| + C(\beta, \epsilon, \sigma) \sum_{m=1}^{\infty} 2^m (\epsilon\sigma e)^{m/\sigma} \cdot \exp\{\epsilon(r + m^{1/\rho})^{2\sigma} + A(r + m^{1/\rho})^{\rho}\}.$$

This series converges for every finite value of z because the m-th root of the m-th term has a limit superior less than or equal to

$$2(\epsilon\sigma e)^{1/\sigma} e^A,$$

and this quantity can be made as small as we please since ϵ is at our disposal. If $\rho = 2$ we have to replace A in the exponent by $A + \epsilon$, but the conclusion is the same. It follows that $G(\delta_z)\cdot f(z)$ exists as an entire function.

If $\rho < 2$ we have $\sigma = 1$ and (5.7) is to be replaced by $u = r + m^{\frac{1}{2}}$ and (5.8) by

$$(5.9)\quad |G(z^2)|\,|f(z)| + C(\epsilon, \sigma) \sum_{m=1}^{\infty} (2\epsilon e)^m \exp\{\epsilon(r + m^{\frac{1}{2}})^2 + A(r + m^{\frac{1}{2}})^{\rho}\}.$$

This series is clearly convergent for every finite value of z. This completes the proof of the theorem.

6. Closer estimates in the conjugate case. We shall now utilize the machinery built up in §§2 and 3 for the purpose of discussing in more detail the case in which $f(z)$ and $G(w)$ are of conjugate orders. We start by proving

THEOREM 6.1. *A necessary and sufficient condition that $G(\delta_z)$ shall be applicable to the class $\mathfrak{F}_{\rho,\alpha}$, ρ fixed > 2, α fixed $(0 < \alpha < \infty)$, is that $G(w) \in \mathfrak{G}_{\sigma,\gamma}$ where $\gamma < \beta$. Here ρ and σ are conjugate orders, α and β conjugate types.*

We shall prove merely the sufficiency here; the necessity of the condition follows from the example in §8.4. By definition

$$G(\delta_z)\cdot f(z) = \sum_{k=0}^{\infty} g_k \delta_z^k f(z).$$

Since $G(w) \in \mathfrak{G}_{\sigma,\gamma}$

$$\limsup_{k\to\infty} \left[\Gamma\left(1 + \frac{k}{\sigma}\right) |g_k|\right]^{1/k} \leqq \gamma^{1/\sigma},$$

while

$$\limsup_{k\to\infty} \left[\frac{|\delta_z^k f(z)|}{\Gamma(1 + k/\sigma)}\right]^{1/k} = [F_\sigma(z; f)]^2 \leqq (2\sigma)^{1/\sigma}(\alpha\rho)^{2/\rho}$$

by formula (3.5). It follows that

$$\limsup_{k\to\infty} |g_k \delta_z^k f(z)|^{1/k} \leqq (2\sigma\gamma)^{1/\sigma}(\alpha\rho)^{2/\rho}.$$

Since $\gamma < \beta$, the conjugate of α, the product is less than 1. Hence $G(\delta_z)\cdot f(z)$ exists as an entire function when the condition of the theorem is satisfied.

If $\rho \leqq 2$, the results are less precise. We start with the case $\rho = 2$.

THEOREM 6.2. *There exists a positive decreasing function $\beta(\alpha)$ such that $G(\delta_z)$ applies to the class $\mathfrak{F}_{2,\alpha}$, α fixed $(0 < \alpha < \infty)$, whenever $G(w) \in \mathfrak{G}_{1,\gamma}$ if and only if $\gamma < \beta(\alpha)$. As $\alpha \to \infty$, $\alpha\beta(\alpha) \to \frac{1}{2}$.*

The existence of such a $\beta(\alpha)$ is proved by a Dedekind cut argument. If $\alpha_1 < \alpha_2$ then $\mathfrak{F}_{2,\alpha_1} \subset \mathfrak{F}_{2,\alpha_2}$, so that if $G(\delta_z)$ applies to $\mathfrak{F}_{2,\alpha_2}$ it also applies to $\mathfrak{F}_{2,\alpha_1}$. Hence the condition $\gamma < \beta(\alpha_2)$ must imply $\gamma < \beta(\alpha_1)$ whence $\beta(\alpha_2) \leqq \beta(\alpha_1)$. For $\beta(\alpha)$ we have the following inequalities

$$\beta_1(\alpha) \leqq \beta(\alpha) \leqq \beta_2(\alpha), \tag{6.1}$$

where

$$\frac{1}{\beta_1(\alpha)} = \sup_f \sup_z [F_1(z;f)]^2, \tag{6.2}$$

$f(z)$ ranging over $\mathfrak{F}_{2,\alpha}$, and

$$\frac{1}{\beta_2(\alpha)} = \sup_f \sup_z [F_1^*(z;f)]^2, \tag{6.3}$$

where

$$F_1^*(z;f) = \lim_{k\to\infty} \left[\frac{|\delta_z^k f(z)|}{\Gamma(1+k)}\right]^{1/(2k)}, \tag{6.4}$$

and $f(z)$ ranges over the subset of functions in $\mathfrak{F}_{2,\alpha}$ for which such a limit exists at some points z. It is quite likely that $\beta_1(\alpha) = \beta_2(\alpha)$, but not being able to get good estimates of either quantity, the author must leave this question unsolved. An example in §8.5 shows that

$$\limsup_{\alpha\to\infty} \alpha\beta_2(\alpha) \leqq \tfrac{1}{2},$$

whence it follows in particular that $\beta(\alpha) \to 0$ as $\alpha \to \infty$. On the other hand, Theorem 2.1 (ii) shows that

$$\liminf_{\alpha\to\infty} \alpha\beta_1(\alpha) \geqq \tfrac{1}{2}.$$

Hence

$$\lim_{\alpha\to\infty} \alpha\beta(\alpha) = \tfrac{1}{2}$$

as claimed above. Formula (2.19) shows that

$$\beta(0) = \lim_{\alpha\to 0} \beta(\alpha) \geqq \frac{1}{2}\left(\frac{3}{e}\right)^{\frac{1}{2}}. \tag{6.5}$$

In case $\rho < 2$ we also have an unsatisfactory situation.

THEOREM 6.3. *If $G(w) \in \mathfrak{G}_{1,\gamma}$ with $\gamma < \beta(0)$, then $G(\delta_z)$ applies to the class $\mathfrak{F}_{2,0}$ and a fortiori to every class $\mathfrak{F}_{\rho,\alpha}$ with $\rho < 2$. There exists a quantity β_0 ($\beta(0) \leqq \beta_0 \leqq 1$) such that $G(\delta_z)$ applies to the class of all polynomials if and only if $G(w) \in \mathfrak{G}_{1,\gamma}$ with $\gamma < \beta_0$.*

The existence of β_0 follows again by a Dedekind cut argument; and since the class of all polynomials is a subset of $\mathfrak{F}_{2,0}$, we have $\beta(0) \leqq \beta_0$. That $\beta_0 \leqq 1$ follows from the example in §8.5 below. It seems quite likely that $\beta_0 = \beta(0)$, but we are unable to prove it.

7. On the order and type of $G(\delta_z)$-transforms. We shall study the relations between orders and types of the three functions $f(z)$, $G(w)$, and $G(\delta_z)\cdot f(z)$. We restrict ourselves to the simplest case.

THEOREM 7.1. *Let $f(z)$ be an entire function of finite order ρ and finite type α. Let $G(w)$ be an entire function of order σ and finite type β. Let ρ' be the conjugate order of ρ, and suppose that $\sigma \leqq \rho'$ and that $\beta = 0$ if $\sigma = \rho'$. Then the entire function $G(\delta_z)\cdot f(z)$ is of order $\mathrm{P} \leqq \max(\rho, 2\sigma)$. If $\mathrm{P} = \rho$ and $\rho > 2\sigma$, the type is at most α; it is at most $\alpha + \beta$ if $\mathrm{P} = \rho = 2\sigma$, and at most β if $\mathrm{P} = 2\sigma$ and $\rho < 2\sigma$. These limits are the best possible.*[20]

We know that $G(\delta_z)\cdot f(z)$ exists as an entire function by Theorem 5.1. In order to get the required estimates we have to modify the analysis which led to formulas (5.8) and (5.9). Suppose first that $\rho > 2\sigma$ and go back to formula (1.20) = (4.2) once more. We choose as contours of integration the circles

$$| t_\nu - t_{\nu+1} | = m^{1/\rho-1} \qquad (\nu = 1, 2, \cdots, m;\ t_{m+1} = z),$$

and in Lemma 4.1 we choose $\sigma = \sigma$, $\beta = \beta$, and

$$u = r + m^{1/\rho}.$$

Recalling that

$$M(r; f) < B(\epsilon) \exp [(\alpha + \epsilon) r^\rho],$$

we see that the difference $G(\delta_z)\cdot f(z) - G(z^2)\cdot f(z)$ is dominated by a quantity $C(\epsilon)$ multiplied into the infinite series

$$\sum_{m=1}^{\infty} 2^m [(\beta + \epsilon)\sigma e]^{m/\sigma} m^{am} \exp \{(\alpha + \epsilon)(r + m^{1/\rho})^\rho + (\beta + \epsilon)(r + m^{1/\rho})^{2\sigma}\},$$

where

$$a = 2\left\{1 - \frac{1}{\rho} - \frac{1}{2\sigma}\right\} \leqq 0.$$

The function $G(z^2)\cdot f(z)$ evidently has all the properties claimed for $G(\delta_z)\cdot f(z)$ in the theorem, i.e., its order is at most $\max(\rho, 2\sigma)$ and its type is at most

[20] See H. Muggli [5], p. 153, for $G(d/dz)$. Muggli does not discuss the type except when $\rho = 1$. See also G. Valiron [10], p. 53.

α, $\alpha + \beta$, and β according as $\rho > 2\sigma$, $\rho = 2\sigma$, and $\rho < 2\sigma$. Moreover, it is only when $\rho = 2\sigma$ that any lowering of order or type can occur for the product. It remains to discuss the infinite series.

We write $\sum = \sum_1 + \sum_2$, where $\sum_1$ contains all terms with $m < r^\rho$ and $\sum_2$ the rest. In $\sum_2$ we have $r + m^{1/\rho} < 2m^{1/\rho}$. Hence

$$\sum_2 < \sum_{m>r^\rho} 2^m[(\beta+\epsilon)\sigma e]^{m/\sigma} m^{am} \exp[(\alpha+\beta+2\epsilon)2^\rho m].$$

We recall that if $a = 0$, $\sigma = \rho'$, and $\beta = 0$; further, ϵ is at our disposal. It follows that $\sum_2 < A(r, \epsilon)$, a finite quantity which tends to zero as $r \to \infty$. In $\sum_1$ we have

$$(r + m^{1/\rho})^\rho < r^\rho + \rho(r + m^{1/\rho})^{\rho-1} m^{1/\rho} < r^\rho + \rho m^{1/\rho}(2r)^{\rho-1}$$

if $\rho > 1$, and

$$(r + m^{1/\rho})^\rho < r^\rho + m$$

if $\rho \leqq 1$. Similarly

$$(r + m^{1/\rho})^{2\sigma} < r^{2\sigma} + 2\sigma m^{1/\rho}(2r)^{2\sigma-1}, \quad \text{or} \quad < r^{2\sigma} + m^{2\sigma/\rho} < r^{2\sigma} + m$$

according as $\sigma > \frac{1}{2}$ or $\sigma \leqq \frac{1}{2}$. Hence

$$\sum_1 < \exp\{(\alpha+\epsilon)r^\rho + (\beta+\epsilon)r^{2\sigma}\} \sum_{m<r^\rho} 2^m[(\beta+\epsilon)\sigma e]^{m/\sigma} m^{am} e^E$$

$$\equiv \exp\{(\alpha+\epsilon)r^\rho + (\beta+\epsilon)r^{2\sigma}\} \textstyle\sum_3.$$

If both $\rho \leqq 1$ and $\sigma \leqq \frac{1}{2}$ we have $E < (\alpha + \beta + 2\epsilon)m$ and $\sum_3$ evidently remains bounded while $r \to \infty$. The most unfavorable case is that in which $\rho > 1$ and $\sigma > \frac{1}{2}$ and we can restrict ourselves to a detailed discussion of this case.

Here we have

$$E < (\alpha+\epsilon)\rho m^{1/\rho}(2r)^{\rho-1} + (\beta+\epsilon)2\sigma m^{1/\rho}(2r)^{2\sigma-1}$$

$$< [(\alpha+\epsilon)\rho + (\beta+\epsilon)2\sigma]m^{1/\rho}(2r)^{\rho-1} \equiv Br^{\rho-1}m^{1/\rho}.$$

$\sum_3$ is then seen to be dominated by an expression of the form

$$\sum_{m<r^\rho} a_m \exp[Br^{\rho-1}m^{1/\rho}],$$

where

$$0 < a_m < A(\Delta)\exp[-\Delta m],$$

Δ being an arbitrarily large fixed quantity. The latter inequality is obvious if $a < 0$, and if $a = 0$ we recall once more that $\beta = 0$ and that ϵ is at our disposal and can be made as small as we please in advance. We note that B is independent of m and r. Now the maximum of

$$\exp\{Br^{\rho-1}m^{1/\rho} - \Delta m\}$$

when m is a continuous variable equals

$$\exp \{(B\rho)^{\rho/(\rho-1)}(\rho - 1)\Delta^{-1/(\rho-1)}r^{\rho}\}.$$

It follows that

$$\textstyle\sum_2 < B(\Delta)r^{\rho} \exp \{C\Delta^{-1/(\rho-1)}r^{\rho}\}.$$

Since Δ is as large as we please, we conclude that

$$\textstyle\sum_1 < D(\epsilon) \exp [(\alpha + 2\epsilon)r^{\rho}].$$

Combining the estimates of $\sum_1$ and $\sum_2$ we see that

$$\textstyle\sum < D_1(\epsilon) \exp [(\alpha + 2\epsilon)r^{\rho}]$$

for every positive ϵ. Consequently, if $\rho > 2\sigma$ the transform $G(\delta_z)\cdot f(z)$ is an entire function of order ρ and type α at most.

If $\rho = 2\sigma$ the same argument gives instead

$$\textstyle\sum < D_1(\epsilon) \exp [(\alpha + \beta + 2\epsilon)r^{\rho}]$$

for every positive ϵ. We conclude that the order of $G(\delta_z)\cdot f(z)$ is at most $\rho = 2\sigma$ and if this is its true order, the type is at most $\alpha + \beta$.

If $\rho < 2\sigma$ we choose the radii of the circles of integration equal to

$$m^{1/(2\sigma)-1}$$

instead. This gives as majorant for the difference $G(\delta_z)\cdot f(z) - G(z^2)\cdot f(z)$ a constant multiple of the series

$$\sum_1^{\infty} 2^m[(\beta + \epsilon)\sigma e]^{m/\sigma} m^{\alpha m} \exp \{(\alpha + \epsilon)(r + m^{1/(2\sigma)})^{\rho} + (\beta + \epsilon)(r + m^{1/(2\sigma)})^{2\sigma}\}.$$

This series is discussed by the same method as above and shows that $G(\delta_z)\cdot f(z)$ now is an entire function of order at most 2σ and that, if this be the true order, the type is at most β.

That the results are the best possible will be proved by examples in §8.6.

Suppose that $\frac{1}{2} \leqq \sigma < \min (\frac{1}{2}\rho, \rho')$ and that $G(w) \in \mathfrak{G}_{\sigma,\beta}$, where β is an arbitrary fixed non-negative real number. Then $G(\delta_z)$ *defines a linear transformation on the class* $\mathfrak{F}_{\rho,\alpha}$ *to itself* by Theorem 7.1. *This transformation appears,* however, *to be neither bounded nor continuous.*[21] An example proving this for the case in which

$$\limsup_{n\to\infty} n^{-\frac{1}{2}} \log | G(2n + 1) | = +\infty$$

will be given in §8.7.

8. Counter examples. In this section we shall give the various examples which will prove our statements in §§2–7 concerning best possible results or

[21] H. Muggli [5], p. 153, showed that $\exp [d^2/dz^2]$ does not define a continuous transformation on the classes to which it applies.

lack of continuity, etc. We shall give as few details as possible since many examples employ the same principle, but the first time the principle is used a fuller treatment will be given.

Many of our examples employ properties of the Hermite functions. We shall list what properties we need here for later reference. We refer the reader to E. Hille [4] for proofs.[22] The function $h_n(z)$ is defined by formula (3) of the introduction. For $r > 0$, $|z| = r$, we have

$$|h_n(z)| \leqq (-i)^n h_n(ir). \tag{8.1}$$

Further,

$$\begin{aligned} 0 &< (-1)^n h_{2n}(ir) - \frac{(2n)!}{n!} \cosh (4n+1)^{\frac{1}{2}} r \\ &< \frac{(2n)!}{n!} \exp [(4n+1)^{\frac{1}{2}} r] \{\exp [\tfrac{1}{6} r^3 (4n+1)^{-\frac{1}{2}}] - 1\}. \end{aligned} \tag{8.2}$$

There is a similar formula for the functions of odd order in which $4n + 1$ is replaced by $4n + 3$, the cosh by sinh, and the factor $(2n)!/n!$ by $2(2n + 1)!/(4n + 3)^{\frac{1}{2}} n!$. The formula

$$(-1)^n \frac{n!}{(2n)!} h_{2n}(ir) = 1 + \frac{n}{2!}(2r)^2 + \frac{n(n-1)}{4!}(2r)^4 + \cdots + \frac{n!}{(2n)!}(2r)^{2n} \tag{8.3}$$

proves that for fixed r the left side is an increasing function of n. A similar result holds for the functions of odd order. If $z = x + iy$, $y > 0$, we have

$$h_n(x + iy) = h_n(iy) \exp [-ix(2n+1)^{\frac{1}{2}}]\{1 + n^{-\frac{1}{2}} \eta_n(x, y)\}, \tag{8.4}$$

where $\eta_n(x, y)$ is bounded for $-1/\epsilon \leqq x \leqq 1/\epsilon$, $\epsilon \leqq y \leqq 1/\epsilon$. Finally we note that if in a Hermitian series the coefficients satisfy the condition $i^n c_n \geqq 0$ for all large n, then the point of intersection of the upper line of convergence with the imaginary axis is a singular point of the function defined by the series.

8.1. Examples for Theorems 2.1 and 3.1. For $\rho > 2$ we shall show the existence of entire functions $f(z)$ of order ρ and type α, such that

$$\lim_{k\to\infty} k^{-1/\sigma} |\delta_z^k f(z)|^{1/k} = \left(\frac{2}{e}\right)^{1/\sigma} (\alpha\rho)^{2/\rho} \tag{8.1.1}$$

for every value of z on a given line $x = x_0$ in the complex plane with the possible exception of the point $z = x_0$. For this purpose we consider the series[23]

$$f(z; x_0, a, b) = \sum_{n=0}^{\infty} (-1)^n \exp \{-a(4n+1)^b + ix_0(4n+1)^{\frac{1}{2}}\} \frac{n!}{(2n)!} h_{2n}(z), \tag{8.1.2}$$

where $x_0 \geqq 0$, $a > 0$, $\frac{1}{2} < b < 1$. We have obviously

$$|f(z; x_0, a, b)| \leqq f(ir; 0, a, b).$$

[22] See, in particular, Theorems 1.1, 1.4, and 5.1.

[23] The case $b = \frac{1}{2}$ is discussed in [4], pp. 895-896.

Let us choose $N = [\frac{1}{24}r^2]$ and break up the series in (8.1.2) into two parts $\sum_1$ and $\sum_2$, where $\sum_1$ contains all the terms with $n \leqq N$ and $\sum_2$ all the rest. Then using (8.2) and (8.3) we get

$$\begin{aligned} \left|\textstyle\sum_1\right| &\leqq \sum_0^N \exp\,[-a(4n+1)^b]\,\frac{n!}{(2n)!}\,|h_{2n}(ir)| \\ &< \frac{N!}{(2N)!}\,|h_{2N}(ir)|\,\sum_0^N \exp\,[-a(4n+1)^b] \\ &< C_1 \exp\,\{(4N+1)^{\frac{1}{2}}r + \tfrac{1}{8}(4N+1)^{-\frac{1}{2}}r^3\} \\ &< C_2 \exp\,\{\tfrac{1}{3}6^{\frac{1}{2}}r^2\}. \end{aligned}$$

Similarly

$$\begin{aligned} \left|\textstyle\sum_2\right| &< \sum_{N+1}^{\infty} \exp\,\{-a(4n+1)^b + r(4n+1)^{\frac{1}{2}} + \tfrac{1}{8}r^3(4n+1)^{-\frac{1}{2}}\} \\ &< \exp\,\{\tfrac{1}{8}6^{\frac{1}{2}}r^2\} \sum_{N+1}^{\infty} \exp\,\{-a(4n+1)^b + r(4n+1)^{\frac{1}{2}}\}. \end{aligned}$$

If r is very large, the exponent in the last formula has a single maximum for $n > N$. It follows by classical arguments that the infinite series is of the same order of magnitude as the integral

$$\int_0^{\infty} \exp\,[-au^{2b} + ru]\,u\,du$$

which by the method of Laplace is found to be less than a constant $A(a, b)$ times

$$r^{(2-b)/(2b-1)} \exp\left\{a(2b-1)\left(\frac{r}{2ab}\right)^{2b/(2b-1)}\right\}.$$

Consequently, $f(z;\, x_0,\, a,\, b)$ is an entire function whose order ρ and type α satisfy the inequalities

$$\rho \leqq \frac{2b}{2b-1} \equiv b', \qquad \alpha \leqq a(2b-1)(2ab)^{-2b/(2b-1)} \equiv a'. \tag{8.1.3}$$

We observe that the quantities which figure on the right sides of these inequalities are the conjugates of b and a respectively in the sense of relations (5.4) and (5.5). We shall prove that equality holds in both places.

Now take $z = z_0 = x_0 + iy_0$, where y_0 is arbitrary but fixed and positive. Then

$$\delta_z^k f(z_0;\, x_0,\, a,\, b) = \sum_{n=0}^{\infty} (-1)^n (4n+1)^k \exp\,\{-a(4n+1)^b + ix_0(4n+1)^{\frac{1}{2}}\} \cdot \frac{n!}{(2n)!}\, h_{2n}(z_0),$$

the absolute value of which exceeds the real part which equals

$$\sum_{n=0}^{\infty} (4n+1)^k \exp[-a(4n+1)^b] \frac{n!}{(2n)!} |h_{2n}(iy_0)| \{1 + n^{-\frac{1}{4}}\Re[\eta_n(x_0, y_0)]\}.$$

Since $\eta_n(x, y)$ is uniformly bounded in $-1/\epsilon \leqq x \leqq 1/\epsilon$, $\epsilon \leqq y \leqq 1/\epsilon$, we can find an integer m such that $n^{-\frac{1}{4}}\Re[\eta_n(x_0, y_0)] > -\frac{1}{2}$ for $n \geqq m$ and any point z_0 in the rectangle mentioned. The terms of the series corresponding to $n < m$ can evidently be neglected for large values of k. The remainder exceeds

$$\begin{aligned}\frac{1}{2}\sum_{m}^{\infty} (4n+1)^k \exp[-a(4n+1)^b] \frac{n!}{(2n)!} |h_{2n}(iy_0)| \\ > \frac{1}{2}\sum_{m}^{\infty} (4n+1)^k \exp[-a(4n+1)^b] \\ > \tfrac{1}{2} \max_n \{(4n+1)^k \exp[-a(4n+1)^b]\} \\ > C\left(\frac{k}{abe}\right)^{k/b}.\end{aligned}$$

It follows that

$$\liminf_{k\to\infty} k^{-1/b} |\delta_z^k f(z_0; x_0, a, b)|^{1/k} \geqq (abe)^{-1/b}. \tag{8.1.4}$$

From this inequality we easily get the required results. We first notice that the inequality can be replaced by the equality

$$\lim_{k\to\infty} k^{-1/b} |\delta_z^k f(z_0; x_0, a, b)|^{1/k} = (abe)^{-1/b}. \tag{8.1.5}$$

Indeed, formula (2.15) is valid not merely for functions of order ρ and type α, but for functions satisfying (3.4), i.e., whose order is at most ρ and type at most α. Using this remark and the estimates (8.1.3), we obtain the inequality

$$\limsup_{k\to\infty} k^{-1/b} |\delta_z^k f(z_0; x_0, a, b)|^{1/k} \leqq (abe)^{-1/b},$$

which combined with (8.1.4) gives (8.1.5).

These formulas show that the functional $F_\lambda(z_0; f)$ is 0, $(abe)^{-1/b}$, or $+\infty$ according as λ is less than, equal to or greater than b. On the other hand, Theorem 3.1 tells us that $F_\lambda(z; f)$ is always 0 for $\lambda \leqq \sigma$, the conjugate of the order of $f(z)$. It follows that $\sigma \leqq b$ and hence that $\rho \geqq b'$. But this is precisely the opposite to the first inequality in (8.1.3). This inequality then must be an equality. From the inequality

$$(abe)^{-1/b} \leqq \left(\frac{2}{e}\right)^{1/b} (\alpha b')^{2/b'},$$

we then get the opposite of the second inequality under (8.1.3) which then also must be an equality.

We have consequently proved that formula (2.15) is the best possible in the sense that equality may actually hold for any preassigned value of z even if the limit superior is replaced by an ordinary limit.[24]

If $\rho = 2$ we can use the function $f(z;\, x_0,\, a,\, 1)$ and proceed as above. We obtain

$$\rho = 2, \qquad \alpha \leqq \tfrac{1}{3}6^{\frac{1}{2}} + \frac{1}{4a}, \tag{8.1.6}$$

$$\liminf_{k\to\infty} k^{-1}\,|\,\delta_z^k f(z_0;\, x_0,\, a,\, 1)\,|^{1/k} \geqq \frac{1}{ae}, \tag{8.1.7}$$

and finally

$$\liminf_{k\to\infty} k^{-1}\,|\,\delta_z^k f(z_0;\, x_0,\, a,\, 1)\,|^{1/k} \geqq \frac{4}{e}\,\alpha - \frac{4}{3}\,6^{\frac{1}{2}}. \tag{8.1.8}$$

This inequality proves that the factor $4/e$ in formula (2.16) cannot be replaced by any smaller quantity.

For $\rho < 2$ or, more precisely, for functions satisfying the condition (2.2) we proved that inequality (2.3) holds. That this inequality is not capable of very considerable improvement is shown by the fact that

$$\limsup_{k\to\infty} k^{-1}\,|\,\delta_z^k 1\,|^{1/k} \geqq \frac{1}{e} \tag{8.1.9}$$

on the lines $y = \pm x$. Indeed, formulas (1.10) and (1.12) show that if $y = \pm x$, then

$$(-1)^\nu \delta_z^{2\nu} 1 \geqq A_{0,2\nu,\nu} > (2\nu)!,$$

and this obviously implies (8.1.9).

8.2. Discontinuities of the functional $F_\lambda(z;f)$. It was stated in §3 that $F_\lambda(z;f)$ may be a discontinuous function of z for a fixed $f(z)$. This is proved by considering the series

$$\varphi(z;\, a,\, b) = \sum_{n=0}^{\infty} (-1)^n \exp\,[-a(4n+3)^b]\, \frac{n!\,(4n+3)^{\frac{1}{2}}}{(2n+1)!}\, h_{2n+1}(z), \tag{8.2.1}$$

where $a > 0$, $\frac{1}{2} \leqq b \leqq 1$. This is a function of the same type as $f(z;\, 0,\, a,\, b)$ of formula (8.1.2) and it can be discussed by the same methods.

When $b = \frac{1}{2}$, the series is convergent in the strip $-a < y < a$ and the points $z = \pm ai$ are singular. It is obvious that

$$F_{\frac{1}{2}}(0;\, \varphi) = 0, \tag{8.2.2}$$

[24] The excluded case in which z_0 is real $\neq 0$ can be handled by a modification of the series. Cf. a similar argument in [4], pp. 896-897.

and it is an easy matter to show that

$$F_{\frac{1}{2}}(iy;\varphi) = \frac{1}{a-|y|} = \frac{1}{R(iy)} \qquad (-a < y < a,\ y \neq 0). \tag{8.2.3}$$

Here $R(z)$ denotes the radius of holomorphism of $\varphi(z)$ at the point z. This example then shows that $F_{\frac{1}{2}}(z;f)$ may be a discontinuous function of z. It also shows, incidentally, that (3.6) is the best possible estimate.

For $\frac{1}{2} < b < 1$ we are dealing with an entire function of order b' and type a' (see formulas (8.1.3)). We have

$$F_b(0;\varphi) = 0, \qquad F_b(iy;\varphi) = a^{-1/(2b)} \qquad (y \neq 0). \tag{8.2.4}$$

Thus $F_b(z;\varphi)$ is discontinuous at $z = 0$. The same result is true when $b = 1$.

8.3. Example for Theorem 4.4. We shall prove that if $G(w)$ is any entire function of order $\frac{1}{2}$ and maximal type, then there exists an entire function $f(z)$, usually of infinite order, such that $G(\delta_z)\cdot f(z)$ does not exist anywhere on the imaginary axis. By assumption we can find a monotone increasing function $\lambda(u)$ tending to infinity with u such that

$$(2k)!\,|\,g_k\,| > [\lambda(k)]^{2k} \tag{8.3.1}$$

for infinitely many values of k. Let us then choose a monotone increasing function $\mu(u)$, tending to infinity with u, such that

$$\lambda(u) \exp\{-\mu(4u^2)\} \to \infty \text{ with } u. \tag{8.3.2}$$

Then form

$$f(z) = \sum_{n=0}^{\infty} (-1)^n \exp\{-(4n+1)^{\frac{1}{2}}\mu(4n+1)\} \frac{n!}{(2n)!} h_{2n}(z). \tag{8.3.3}$$

It is easily seen that this is an entire function which is ordinarily of infinite order. Further

$$\delta_z^k f(iy) > \delta_z^k f(0) > C(2k)^{2k} \exp\{-2k\mu(4k^2)\}. \tag{8.3.4}$$

It follows that the series

$$\sum_0^{\infty} g_k \delta_z^k f(z)$$

cannot converge anywhere on the imaginary axis since there is a subsequence of the terms tending to $+\infty$. On the other hand, it may very well happen that the series

$$\sum_{n=0}^{\infty} (-1)^n \exp\{-(4n+1)^{\frac{1}{2}}\mu(4n+1)\} G(4n+1) \frac{n!}{(2n)!} h_{2n}(z), \tag{8.3.5}$$

obtained by termwise performance of the operation $G(\delta_z)$ on (8.3.3), converges for all values of z. This depends exclusively upon the asymptotic behavior of $G(4n+1)$ for large values of n.

8.4. Example for Theorems 5.1 and 6.1. The following example really refers to Theorem 6.1 but it is also a counter example for that part of Theorem 5.1 in which $\rho > 2$. We suppose then that $\rho > 2$ and let ρ and σ be conjugate orders and α and β conjugate types. Corresponding to a given entire function $G(w)$ of order σ and type β we shall show the existence of an entire function $f(z)$ of order ρ and type α, such that $G(\delta_z)\cdot f(z)$ does not exist. By assumption

$$(8.4.1) \qquad |g_k| = \frac{\beta^{k/\sigma}}{\Gamma(k/\sigma)}\lambda_k \quad \text{where} \quad \frac{1}{k}\log \lambda_k \to 0$$

as $k \to \infty$. The unfavorable case for us is that in which $\lim\inf \lambda_k = 0$. It is then possible to find a steadily decreasing function $\lambda(u)$ such that $\lambda(u) \to 0$, $u^{-1} \log \lambda(u) \to 0$ as $u \to \infty$ but $\lambda_k \geqq \lambda(k)$ for infinitely many values of k. We then form the following modification of our first counter example in formula (8.1.2):

$$(8.4.2) \qquad f(z) = \sum_{n=0}^{\infty} (-1)^n \frac{\exp[-\beta(4n+1)^\sigma]}{\lambda[\sigma\beta(4n+1)^\sigma]} \frac{n!}{(2n)!} h_{2n}(z).$$

It is not difficult to show that this is an entire function of order ρ and type α. On the imaginary axis

$$(8.4.3) \qquad \delta_z^k f(iy) > \frac{C}{\lambda(k)}\left(\frac{k}{\beta\sigma e}\right)^{k/\sigma},$$

whence it follows that the series

$$(8.4.4) \qquad \sum_{k=0}^{\infty} g_k \delta_z^k f(z)$$

diverges everywhere on the imaginary axis, because it contains infinitely many terms the absolute values of which exceed a positive constant.

The converse problem: *Given any function $f(z)$ of order ρ and type σ, construct an entire function $G(w)$ of conjugate order σ and conjugate type β, such that $G(\delta_z)\cdot f(z)$ does not exist,* appears to be much more difficult. It would be easily solved if it were known that for every entire function of order ρ and type α there exists at least one point z_0 where

$$F_\sigma(z_0\,; f) = (2\sigma)^{1/(2\sigma)}(\alpha\rho)^{1/\rho}.$$

Whether or not this is actually true is one of the open questions which we cannot answer.

The function defined by (8.4.2) shows that the condition of Theorem 6.1 is necessary for the truth of that theorem. It also shows that the condition of Theorem 5.1 is necessary when $\rho > 2$. If we set $\sigma = 1$ in (8.4.2), we obtain an entire function of order 2 for which the series (8.4.4) diverges on the imaginary axis. Thus the condition of Theorem 5.1 is also necessary when $\rho = 2$. That it is not necessary when $\rho < 2$ follows already from Theorem 6.3.

8.5. Examples for Theorems 6.2 and 6.3. We have first to prove formula (6.5). For this purpose we consider the function $f(z; 0, a, 1)$, a special case of the function defined by formula (8.1.2). Here we can sharpen (8.1.6) and determine the exact type of the function. We have[25]

$$f(iy; 0, a, 1) = \sum_{n=0}^{\infty} (-1)^n \frac{n!}{(2n)!} h_{2n}(iy) e^{-(4n+1)a}$$

$$> e^{-a} \sum_{n=0}^{\infty} (-1)^n \frac{1}{4^n n!} h_{2n}(iy) e^{-4na}$$

$$= \left\{\frac{\pi}{2 \sinh 2a}\right\}^{\frac{1}{2}} \exp\left[\tfrac{1}{2} \coth 2a\, y^2\right].$$

In the opposite direction we get, for instance, by a suitable use of Cauchy's inequality, that

$$f(iy; 0, a, 1) < C(a) \mid y \mid \exp\left[\tfrac{1}{2} \coth 2a\, y^2\right],$$

where the exact value of $C(a)$ is immaterial. It follows that

$$\alpha = \tfrac{1}{2} \coth 2a. \tag{8.5.1}$$

Further, a more elaborate analysis shows that we can sharpen (8.1.7) in the present case to

$$\lim_{k\to\infty} k^{-1} \mid \delta_z^k f(iy; 0, a, 1) \mid^{1/k} = \frac{1}{ae}. \tag{8.5.2}$$

Thus the functional $F_1^*(z; f)$ defined by formula (6.4) has a sense for $f(z; 0, a, 1)$ at least on the imaginary axis. By virtue of (6.3) this leads to the simple inequality

$$\beta_2(\alpha) < a = \tfrac{1}{4} \log \frac{2\alpha + 1}{2\alpha - 1}, \quad \alpha > \tfrac{1}{2}, \tag{8.5.3}$$

of which (6.5) is an immediate consequence. We notice that $\beta(\alpha)$ also satisfies the inequality (8.5.3).

We next have to prove that the quantity β_0 introduced in Theorem 6.3 is less than or equal to one. This follows from the fact that

$$\exp(-\beta\delta_z)\cdot 1$$

does not exist at the origin for any $\beta \geqq 1$ since $(-1)^\nu(\delta_z^{2\nu} 1)_{z=0} > (2\nu)!$.

8.6. Examples for Theorem 7.1. We shall show that if $\rho > 2\sigma$, the order and type of $G(\delta_z)\cdot f(z)$ may actually coincide with those of $f(z)$. We restrict ourselves to the case $\rho \geqq 2$. The function $f(z; 0, a, b)$ of formula (8.1.2) is an

[25] The second equality follows from the Abel-Hermite kernel also known as Mehler's generating function for Hermite polynomials.

entire function of order b' and type a' (see formula (8.1.3)). We take $G(w) = E_{1/\sigma}(w)$, the Mittag-Leffler E-function, where $\sigma < b$. Then

$$(8.6.1) \quad \begin{aligned} E_{1/\sigma}(\delta_z)\cdot f(z; 0, a, b) &= \sum_{n=0}^{\infty} (-1)^n E_{1/\sigma}(4n+1) \\ &\quad \cdot \exp\,[-a(4n+1)^b]\, \frac{n!}{(2n)!}\, h_{2n}(z), \end{aligned}$$

and with the aid of the methods of §8.1 it is a simple matter to prove that this is also an entire function of order b' and type a' as long as $\sigma < b$. If $\sigma = b$ but $a > 1$, the order is still b' but the type increases to infinity as $a \to 1$. The transform does not exist for $a \leqq 1$.

Finally, we shall give an example to show that if $\rho < 2\sigma$, the transform may be of order 2σ and have the same type as $G(w)$. We take $f(z) = 1$ and

$$(8.6.2) \qquad G(w) = \sum_{k=0}^{\infty} \frac{(aw)^{4k}}{\Gamma(1 + 4k/\sigma)}.$$

If $a > 0$, this is an entire function of order σ and type a^σ.[26] We assume $\sigma \leqq 1$, and apply formula (1.15), obtaining

$$(8.6.3) \qquad G(\delta_z)\cdot 1 = \sum_{k=0}^{\infty} \frac{a^{4k}}{\Gamma(1 + 4k/\sigma)} \sum_{m=0}^{2k} (-1)^m A_{0,4k,m} z^{8k-4m}.$$

Here the operator $G(\delta_z)$ satisfies condition (1.16), so we can use the remarks made in connection with this condition. As the class $\mathfrak{F}$ we can take either $\mathfrak{F}_{2,0}$ or simply $\mathfrak{P}$, the class of all polynomials. In either case the class is left invariant by transformations which affect merely the arguments of the derivatives at the origin, leaving their absolute values unchanged. If $\sigma < 1$ it is known that the operator $G(\delta_z)$ applies to the class $\mathfrak{F}$; if $\sigma = 1$ the assumption of applicability imposes a condition on a. If $G(\delta_z)$ does apply, we know that the series (8.6.3) is absolutely convergent for all z. Hence we have

$$(8.6.4) \qquad G(\delta_z)\cdot 1 = \sum_{\nu=0}^{\infty} (-1)^\nu z^{4\nu} \sum_{2k \geqq \nu} \frac{A_{0,4k,2k-\nu}}{\Gamma(1 + 4k/\sigma)}\, a^{4k}.$$

This expression reaches its maximum on the lines $y = \pm x$. Substituting $z = \omega r$, where ω is a primitive eighth root of unity, and noting that the A's are positive integers, we find readily that

$$(8.6.5) \qquad G(\delta_z)\cdot 1 \geqq G(r^2) > C(\sigma) \exp\,(a^\sigma r^{2\sigma}).$$

It follows that $G(\delta_z)\cdot 1$ is an entire function whose order is exactly 2σ and whose type is a^σ. This conclusion is valid when $\sigma < 1$. If $\sigma = 1$ the existence of the transform is ensured only for sufficiently small values of a. For such values the order of the transform equals $2\sigma = 2$, but the type exceeds a and tends to

[26] See formula (4.6). $G(w)$ is evidently a linear combination of Mittag-Leffler E-functions.

infinity as a approaches the maximal value beyond which the transform does not exist.

8.7. Discontinuity of $G(\delta_z)$ on $\mathfrak{F}_{\rho,\alpha}$. Let us suppose that $G(w)$ is an entire function of order σ ($\frac{1}{2} \leqq \sigma \leqq 1$) such that

$$(8.7.1) \qquad \limsup_{n\to\infty} n^{-\frac{1}{2}} \log |G(2n+1)| = +\infty.$$

If $2\sigma < \rho < \sigma'$, the conjugate of σ, the operator $G(\delta_z)$ defines a linear transformation on $\mathfrak{F}_{\rho,\alpha}$ to itself. Suppose in addition that $\rho \geqq 2$. I say that *the transformation $G(\delta_z)$ is not continuous and a fortiori not bounded on $\mathfrak{F}_{\rho,\alpha}$.*

By virtue of (8.7.1) we can find a positive monotone increasing function $\lambda(u)$ tending to infinity with u, such that

$$(8.7.2) \qquad |G(2n+1)| > \exp[(2n+1)^{\frac{1}{2}}\lambda(n)]$$

for infinitely many values of n. We can assume, without restriction of the generality, that there are infinitely many even values of n for which (8.7.2) is true. Put $\mu(u) = [\lambda(u)]^{\frac{1}{2}}$, and define

$$(8.7.3) \qquad f_n(z) = \frac{n!}{(2n)!}\exp[-(4n+1)^{\frac{1}{2}}\mu(2n)]\,h_{2n}(z).$$

Since $\rho \geqq 2$, these functions belong to $\mathfrak{F}_{\rho,\alpha}$,[27] and by virtue of formula (8.2) and the properties of $\mu(u)$ the sequence $\{f_n(z)\}$ converges to zero as $n \to \infty$, uniformly in any fixed circle $|z| \leqq R$. On the other hand,

$$(8.7.4) \qquad G(\delta_z)\cdot f_n(z) = G(4n+1)f_n(z)$$

obviously does not converge to zero anywhere and $\limsup |G(\delta_z)\cdot f_n(z)| = \infty$ everywhere outside of the real axis. This proves that $G(\delta_z)$ is not continuous at $f = 0$ and hence nowhere in $\mathfrak{F}_{\rho,\alpha}$.

9. Additional comments on $G(\delta_z)$ and related operators. Let us first point out that the investigation given here of the operator $G(\delta_z)$ also extends to the more general operator

$$(9.1) \qquad G(\delta_z, z, m) = \sum_{\nu=0}^{m} a_\nu z^\nu G_\nu(\delta_z),$$

where the a's are given constants and the $G_\nu(w)$ given entire functions. This operator is equivalent to

$$(9.2) \qquad \sum_{k=0}^{\infty} P_k(z)\delta_z^k,$$

where $\{P_k(z)\}$ is a given sequence of polynomials of degree $\leqq m$.

[27] If $\rho = 2$, this requires $\alpha \geqq \frac{1}{2}$.

We find that the operator $G(\delta_z, z, m)$ applies to all analytic functions if and only if all $G_\nu(w) \in \mathfrak{G}_{1,0}$. The operator applies to all functions of $\mathfrak{F}_{\rho,\alpha}$, ρ fixed $\geqq 2$, if and only if all $G_\nu(w) \in \mathfrak{G}_{\sigma,0}$, where σ is the conjugate of ρ, and so on.

Our remaining remarks are devoted to a comparison between the two operators $G(d/dz)$ and $G(\delta_z)$. The former operator is fairly well known, having been the object of much research in the past. There is much similarity between the two theories and we have called attention to such features in several places above. But there are also considerable differences. The applicability questions are much easier to solve for the operator $G(d/dz)$ than for $G(\delta_z)$. This is of course essentially because of the fact that it is much easier to discuss the rate of growth of $f^{(k)}(z)$ than of $\delta_z^k f(z)$. But the difference is not merely a difference in degree of accessibility to customary analytical technique. This would not be so interesting if the general situation were fundamentally the same in both cases. Actually there seem to exist differences of more profound nature.

One such difference is reflected in the different character of the order relations which govern the applicability of these operators to entire functions. These relations are[28]

$$\text{(9.3)} \qquad \frac{1}{\rho} + \frac{1}{\sigma} = 1 \quad \text{and} \quad \frac{1}{\rho} + \frac{1}{2\sigma} = 1$$

for d/dz and δ_z respectively. The difference in the coefficients of the formulas is immaterial. More essential is the fact that the first formula is valid whenever it has a sense, i.e., for $1 \leqq \sigma \leqq \infty$, while the second one holds only for $\frac{1}{2} \leqq \sigma \leqq 1$. Thus there is always a class $\mathfrak{F}_{\rho,\alpha}$ of entire functions $f(z)$ to which the operator $G(d/dz)$ applies when $G(w)$ is an entire function. Not so with $G(\delta_z)$. Here we can find a class $\mathfrak{F}_{\rho,\alpha}$ to which it applies only if $G(w) \in \mathfrak{G}_{\sigma,\beta}$ for $0 \leqq \sigma \leqq 1$ and not for any $\sigma > 1$. Moreover, if $G(w)$ is merely holomorphic in a finite circle $|w| < R$, the operator $G(d/dz)$ always applies to the class $\mathfrak{F}_{1,0}$, while the class of functions to which $G(\delta_z)$ applies appears to be highly special and certainly does not contain any class $\mathfrak{F}_{\rho,\alpha}$ as a core. Thus the phenomenon of a critical ρ-order and a maximal σ-order affects the operator $G(\delta_z)$ profoundly and there is no correspondence in the theory of $G(d/dz)$.

This phenomenon is intimately connected with the difference in behavior between the two basic functionals

$$\text{(9.4)} \qquad \Phi_\sigma(z; f) = \limsup_{k\to\infty} \left| \frac{f^{(k)}(z)}{\Gamma(1 + k/\sigma)} \right|^{1/k},$$

$$\text{(9.5)} \qquad F_\sigma(z; f) = \limsup_{k\to\infty} \left| \frac{\delta_z^k f(z)}{\Gamma(1 + k/\sigma)} \right|^{1/(2k)}.$$

We observed in §3 that if $f(z)$ is holomorphic at a point z whose distance from the nearest singularity is $R(z)$, then

$$\Phi_1(z; f) = \frac{1}{R(z)}, \qquad F_1(z; f) \leqq \frac{1}{R(z)},$$

[28] See H. Muggli [5], p. 152, for the first operator.

and the former functional is a continuous function of z within the domain of holomorphism while the latter need not be. If $f(z) \in \mathfrak{F}_{\rho,\alpha}$ $(1 < \rho)$, and σ is determined from the first formula under (9.3), then $\Phi_\sigma(z;f)$ exists and is a function of ρ and α, independent of z and $f(z)$. On the other hand, if $f(z) \in \mathfrak{F}_{\rho,\alpha}$ $(2 < \rho)$, and σ is determined from the second formula under (9.3), then $F_\sigma(z;f)$ exists as a finite quantity but depends upon $f(z)$ and may be a discontinuous function of z. No matter what value σ has $(0 < \sigma \leqq \infty)$, there is always a class of entire functions for which $\Phi_\sigma(z;f)$ is bounded everywhere. In particular, for $\sigma = \infty$, $\Phi_\infty(z;f) \leqq \alpha$ whenever $f(z) \in \mathfrak{F}_{1,\alpha}$, i.e., for every function of *exponential type* in Pólya's terminology, the type being less than or equal to α.

$F_\sigma(z;f)$ shows an entirely different behavior. In particular, the class of entire functions for which $F_\infty(z;f) \leqq \alpha$ seems to be quite special and connected with the more intricate part of the theory of the Hermite-Weber equation. Any solution of the equation

$$w'' + (2\kappa + 1 - z^2)w = 0, \qquad |2\kappa + 1| \leqq \alpha^2,$$

belongs to this class and other functions of the class can be generated by customary analytical devices from such solutions.

The determination of the class of functions for which $F_\infty(z;f) \leqq \alpha$ would seem to be of some importance. Indeed, the classical theory of the differential equation $G(d/dz)\cdot W = F(z)$ has been largely concerned with the class of functions for which $\Phi_\infty(z;f) \leqq \alpha$, i.e., the class $\mathfrak{F}_{1,\alpha}$. It is essentially this class which serves as the basis of the investigations of R. D. Carmichael, H. T. Davis, E. Hilb, H. von Koch, O. Perron, G. Pólya, and I. M. Sheffer, to mention only a few.[29] Moreover, F. Schürer [8] has developed a general theory of L-operations, satisfying certain postulates, which includes also the theory of the functional equation $G(L)\cdot W = F(z)$. He determines all solutions of this equation within the class of functions for which

$$\limsup_{k\to\infty} |L^k f(z)|^{1/k} \leqq q. \tag{9.6}$$

It is easy to formulate conditions under which Schürer's postulates are satisfied for $L = \delta_z$, but as long as the class of functions for which $F_\infty(z;f) \leqq q^2$ is not well defined, the existence theorems given by the Schürer theory are not of much interest in the present case.

Let us finally call attention to one more feature that makes for a distinct difference between the operators $G(d/dz)$ and $G(\delta_z)$. It goes back to the characteristic functions of the operators d/dz and δ_z. Since

$$\frac{d}{dz} e^{\lambda z} = \lambda e^{\lambda z}, \tag{9.7}$$

we can say that every complex number λ is a characteristic value of the operator d/dz with $e^{\lambda z}$ as the corresponding characteristic function. The growth prop-

[29] See R. D. Carmichael [1] and H. T. Davis [2] for further references and an outline of the results obtained by these writers.

erties of the characteristic functions are determined entirely by the characteristic value in question and the operator d/dz does not single out any particular value of λ or any particular direction in the z-plane.

The operator δ_z shows a different behavior. The equation[30]

$$\delta_z f(z) = (2\kappa + 1)f(z) \tag{9.8}$$

is satisfied by the Hermite-Weber functions of order κ. Again all complex numbers are characteristic values. But now all odd integers are exceptional in the sense that some of the corresponding solutions have exceptional growth properties, namely, $h_n(z)$ when $\kappa = n$ and $h_n(iz)$ when $\kappa = -n - 1$. Moreover, the growth properties are governed by κ only when it comes to the fine structure. In the first approximation it is much more important to know which particular solution is considered than the value of κ. The lines $y = \pm x$ divide the plane into four sectors in each of which there is a solution of (9.8) which tends exponentially to zero as $z \to \infty$ regardless of the value of κ which contributes only a minor correction. As a rule these four subdominant solutions are pair-wise linearly independent and dependence occurs if and only if κ is an integer.

Thus the operator δ_z is strongly oriented in the complex z-plane and gives a certain preference to the odd integers among the characteristic values. This orientation and preference shows itself in many ways in the properties of the operator $G(\delta_z)$ and becomes particularly important in the theory of the differential equation $G(\delta_z)\cdot W = F(z)$ which will be taken up for study in a later paper.

[The prototype of Theorem 4.2 for the operator $G(d/dz)$ is due to J. F. Ritt [7], pp. 34–35. I use this opportunity to amend some statements in my paper [4]. In footnote 4 credit for the first application of finite order differential operators to analytic continuation should not have been assigned to H. Cramér as the publication of Ritt's thesis preceded that of Cramér's by five months. Ritt also has the honor of having proved the first general gap theorem for Dirichlet series, preceding both Carlson-Landau and Szász, and this should have been pointed out in footnote 28. I apologize for these unintentional oversights. It should be observed, however, that my Theorems 4.3 and 5.7 are analogues of theorems due to Cramér and Carlson-Landau-Szász and not of related theorems due to Ritt. My methods, aside from the basic differential operator approach, have very little in common with those of Ritt. I expect to make extensive use of Ritt's methods in later papers, however. Added November 15, 1940.]

Bibliography

1. R. D. Carmichael, *Linear differential equations of infinite order*, Bulletin Amer. Math. Soc., vol. 42(1936), pp. 193-218.
2. H. T. Davis, *The Theory of Linear Operators from the Standpoint of Differential Equations of Infinite Order*, Bloomington, Ind., 1936, xiv + 628 pp.
3. P. Flamant, *La notion de continuité dans l'étude des transmutations distributives des fonctions d'une variable complexe et ses applications*, Bulletin des Sci. Math., (2), vol. 52(1928), pp. 26-48, 77-96, 104-128.

[30] For the properties of the solutions of the Weber-Hermite equation see E. T. Whittaker and G. N. Watson [11], Chapter XVI. They refer to them as the *functions of the parabolic cylinder*.

4. E. Hille, *Contributions to the theory of Hermitian series*, this Journal, vol. 5(1939), pp. 875–936.
5. H. Muggli, *Differentialgleichungen unendlich hoher Ordnung mit konstanten Koeffizienten*, Commentarii Math. Helvetici, vol. 11(1938), pp. 151–179.
6. G. Pólya, *Untersuchungen über Lücken und Singularitäten von Potenzreihen*, Math. Zeitschrift, vol. 29(1929), pp. 549–640.
7. J. F. Ritt, *On a general class of linear homogeneous differential equations of infinite order with constant coefficients*, Trans. Amer. Math. Soc., vol. 18(1917), pp. 27–49.
8. F. Schürer, *Eine gemeinsame Methode zur Behandlung gewisser Funktionalgleichungsprobleme*, Berichte d. Sächsichen Ges. d. Wiss. zu Leipzig, Math.-Phys. Kl., vol. 70(1918), pp. 185–240.
9. G. Valiron, *Lectures on the General Theory of Integral Function*, Toulouse, 1923, xi + 208 pp.
10. G. Valiron, *Sur les solutions des équations différentielles linéaires d'ordre infini et à coefficients constants*, Annales de l'École Norm. Sup., (3), vol. 46(1929), pp. 25–53.
11. E. T. Whittaker and G. N. Watson, *Modern Analysis*, 4th edition, Cambridge University Press, 1927, x + 608 pp.

Yale University.

FUNCTIONS WITH POSITIVE DIFFERENCES

By R. P. Boas, Jr. and D. V. Widder

This note had its origin in an effort to make accessible in the literature the proof of a lemma used by Widder[1] in the study of the bilateral Laplace transform. The result in question was that a continuous function which has all differences of even order non-negative in an interval is necessarily analytic there. That it was true was fairly evident from earlier work of S. Bernstein,[2] though the details required for its demonstration seemed not to be available. Following a very natural inductive method we were easily able to supply a proof. But we soon saw that our method would prove a great deal more. We were able to show in fact that a continuous function which has a single difference, say of order k, of constant sign throughout an interval is of class C^{k-2} there, has right-hand and left-hand derivatives of order $k - 1$, and is convex or concave in pieces as if it had a k-th derivative of constant sign. This proved, it is clear that the function of the lemma has derivatives of all orders. A theorem of S. Bernstein[2] then guarantees its analyticity. But a new proof of this and related facts will be given by Boas in a separate note.

After completing our proof we discovered that the result had been proved earlier by T. Popoviciu.[3] Since our method is simpler and more direct for the purpose in hand, we believe its publication will be of value. We point out one main difference in the two methods of attack. Popoviciu makes his discussion depend on divided differences involving unequally spaced points, whereas we deal entirely with differences involving only equally spaced points.

Definition 1.

$$\Delta_\delta^0 f(x) = f(x),$$
$$\Delta_\delta^1 f(x) = \Delta_\delta f(x) = f(x + \delta) - f(x),$$
$$\Delta_\delta^k f(x) = \Delta_\delta^{k-1} f(x + \delta) - \Delta_\delta^{k-1} f(x) \qquad (k = 2, 3, \cdots).$$

It is easy to establish by induction the useful formula

$$\Delta_\delta^k f(x) = \sum_{i=0}^{k} (-1)^{k-i} \binom{k}{i} f(x + i\delta).$$

Received September 21, 1940; presented to the American Mathematical Society, September 10, 1940.

[1] D. V. Widder, *Necessary and sufficient conditions for the representation of a function by a doubly infinite Laplace integral*, Bulletin of the American Mathematical Society, vol. 40(1934), pp. 321–326.

[2] S. Bernstein, *Leçons sur les propriétés extrémales et la meilleure approximation des fonctions analytiques d'une variable réelle*, Paris, 1926, pp. 190-197.

[3] T. Popoviciu, *Sur l'approximation des fonctions convexes d'ordre supérieur*, Mathematica (Cluj), vol. 8(1934), pp. 1-85, especially pp. 54–58.

We remark that Δ operators are evidently permutable: $\Delta_\delta^k[\Delta_\epsilon^l f(x)] = \Delta_\delta^k[\Delta_\epsilon^l f(x)]$; hence we regularly omit the brackets from such expressions.

DEFINITION 2. *A real function $f(x)$ satisfies condition H_k $(k = 1, 2, \cdots)$ in $a < x < b$ provided that* (i) *$f(x)$ is continuous in $a < x < b$ and*

(ii) $$\Delta_\delta^k f(x) \geqq 0 \qquad (a < x < x + k\delta < b).$$

THEOREM. *If $f(x)$ satisfies H_k $(k > 2)$ in $a < x < b$, $f^{(k-2)}(x)$ exists in $a < x < b$, is continuous and convex, and has non-decreasing right-hand and left-hand derivatives.*

We prove the theorem by a series of lemmas, some of which state well-known properties of convex functions; the proofs of these are included for completeness.

LEMMA 1. *If $f(x)$ satisfies H_k $(k \geqq 2)$ in $a < x < b$, then for any k positive numbers $\delta_1, \delta_2, \cdots, \delta_k$*

$$\Delta_{\delta_1}\Delta_{\delta_2} \cdots \Delta_{\delta_k} f(x) \geqq 0$$

provided that $a < x < x + \delta_1 + \delta_2 + \cdots + \delta_k < b$.

The proof is by induction on k. If $h > 0$, n is a positive integer, and $x + h/n + (k - 1)h < b$,

$$\Delta_h f(x) = \sum_{i=0}^{n-1} \Delta_{h/n} f(x + ih/n),$$

$$\Delta_h^{k-1} f(x) = \sum_{i_1=0}^{n-1} \cdots \sum_{i_{k-1}=0}^{n-1} \Delta_{h/n}^{k-1} f(x + [i_1 + \cdots + i_{k-1}]h/n),$$

(1) $$\Delta_{h/n}\Delta_h^{k-1} f(x) = \sum_{i_1=0}^{n-1} \cdots \sum_{i_{k-1}=0}^{n-1} \Delta_{h/n}^k f(x + [i_1 + \cdots + i_{k-1}]h/n) \geqq 0.$$

If $x + mh/n + (k - 1)h < b$, we apply (1) to $x, x + h/n, \cdots, x + (m - 1)h/n$, obtaining

(2) $$\Delta_h^{k-1} f(x) \leqq \Delta_h^{k-1} f(x + h/n) \leqq \cdots \leqq \Delta_h^{k-1} f(x + mh/n).$$

If $x + \delta_1 + (k - 1)h < b$, we choose a sequence of rational numbers m/n such that $x + mh/n + (k - 1)h < b$ and $m/n \to \delta_1/h$; since $f(x)$ is continuous,

$$\Delta_h^{k-1} f(x + mh/n) \to \Delta_h^{k-1} f(x + \delta_1),$$

and (2) yields

$$\Delta_h^{k-1} f(x) \leqq \Delta_h^{k-1} f(x + \delta_1),$$

or

(3) $$\Delta_h^{k-1}\Delta_{\delta_1} f(x) \geqq 0.$$

If $k = 2$, (3) is the desired conclusion. If $k > 2$, we suppose the lemma established for $k - 1$ and prove it for k. Inequality (3) shows that $\Delta_{\delta_1} f(x)$, with fixed δ_1 $(0 < \delta_1 < b - a)$, satisfies H_{k-1} in $a < x < b - \delta_1$. Then

$$\Delta_{\delta_2}\Delta_{\delta_3} \cdots \Delta_{\delta_k}\Delta_{\delta_1} f(x) \geqq 0,$$

provided that $a < x + \delta_1 + \delta_2 + \cdots + \delta_k < b$; since δ_1 is arbitrary, the conclusion of Lemma 1 follows for k.

LEMMA 2. *If $f(x)$ satisfies H_k $(k \geqq 2)$ in (a, b), $\Delta_\epsilon^{k-1}f(x)$ and $\Delta_\epsilon^{k-1}f(x - \epsilon)$ are non-decreasing functions of x in $a < x < b - (k - 1)\epsilon$ and $a + \epsilon < x < b - (k - 2)\epsilon$ respectively.*

If $a < y < z < b - (k - 1)\epsilon$, write $\delta = z - y$. In Lemma 1, take $\delta_1 = \delta$, $\delta_2 = \cdots = \delta_k = \epsilon$. Then $\Delta_\delta\Delta_\epsilon^{k-1}f(y) \geqq 0$, or $\Delta_\epsilon^{k-1}f(z) \geqq \Delta_\epsilon^{k-1}f(y)$; this proves the first part.

If $a + \epsilon < y < z < b - (k - 2)\epsilon$, write $\delta = z - y$, $\delta_1 = \delta$, $\delta_2 = \cdots = \delta_k = \epsilon$. Then $\Delta_\delta\Delta_\epsilon^{k-1}f(y - \epsilon) \geqq 0$, or $\Delta_\epsilon^{k-1}f(z - \epsilon) \geqq \Delta_\epsilon^{k-1}f(y - \epsilon)$; this proves the second part.

Common hypothesis of Lemmas 3–6: $f(x)$ satisfies H_2 in $a < x < b$.

LEMMA 3. *If $a < x < b$, $h^{-1}\Delta_h f(x)$ is a non-decreasing function of h in $a - x < h < b - x$.*

Remark. Since $\Delta_{-h}f(x) = -\Delta_h f(x - h)$, we can also say that $h^{-1}\Delta_h f(x - h)$ is a non-increasing function of h in $0 < h < x - a$.

Suppose $0 < \epsilon < \delta$, $x + \delta < b$. Then

$$\Delta_{\delta/n}^2 f(x) \geqq 0,\ \Delta_{\delta/n}^2 f(x + \delta/n) \geqq 0, \cdots, \Delta_{\delta/n}^2 f[x + (n - 2)\delta/n] \geqq 0,$$

$$\Delta_{\delta/n}f(x) \leqq \Delta_{\delta/n}f(x + \delta/n) \leqq \cdots \leqq \Delta_{\delta/n}f[x + (n - 1)\delta/n]. \tag{4}$$

If $0 < m < n$, the average of the first m terms of (4) does not exceed the average of the first n terms. Hence

$$\frac{f(x + m\delta/n) - f(x)}{m\delta/n} \leqq \frac{f(x + \delta) - f(x)}{\delta}. \tag{5}$$

Choose a sequence of rational fractions m/n such that $m/n \to \epsilon/\delta$; from (5) we deduce

$$\frac{1}{\epsilon}\,\Delta_\epsilon f(x) \leqq \frac{1}{\delta}\,\Delta_\delta f(x) \qquad (0 < \epsilon < \delta).$$

This proves the lemma for positive h. We have similarly, if $0 < \epsilon < \delta$ and $x - \delta > a$,

$$\Delta_{\delta/n}f(x - \delta) \leqq \Delta_{\delta/n}f[x - (n - 1)\delta/n] \leqq \cdots \leqq \Delta_{\delta/n}f(x - \delta/n),$$

$$\frac{\Delta_\delta f(x - \delta)}{\delta} \leqq \frac{\Delta_{m\delta/n}f(x - m\delta/n)}{m\delta/n},$$

$$\frac{\Delta_\delta f(x - \delta)}{\delta} \leqq \frac{\Delta_\epsilon f(x - \epsilon)}{\epsilon} \qquad (0 < \epsilon < \delta),$$

$$\frac{\Delta_{-\delta}f(x)}{-\delta} \leqq \frac{\Delta_{-\epsilon}f(x)}{-\epsilon} \qquad (-\delta < -\epsilon < 0).$$

DEFINITION 3.

$$f'_+(x) = \lim_{\delta \to 0+} \frac{\Delta_\delta f(x)}{\delta},$$

$$f'_-(x) = \lim_{\delta \to 0-} \frac{\Delta_\delta f(x)}{\delta} = \lim_{\delta \to 0+} \frac{\Delta_\delta f(x - \delta)}{\delta}.$$

LEMMA 4. *$f'_+(x)$ and $f'_-(x)$ exist and are finite and non-decreasing in $a < x < b$.*

Lemma 3 shows that $f'_+(x)$ and $f'_-(x)$ exist (finite or infinite). If $0 < \delta < \epsilon$, and $a < z - \epsilon < x - \epsilon < x < x + \epsilon < y + \epsilon < b$, we have by Lemmas 2 and 3 (using the remark after Lemma 3),

$$\text{(6)} \qquad \frac{\Delta_\epsilon f(z - \epsilon)}{\epsilon} \leqq \frac{\Delta_\epsilon f(x - \epsilon)}{\epsilon} \leqq \frac{\Delta_\delta f(x - \delta)}{\delta} \leqq \frac{\Delta_\delta f(x)}{\delta} \leqq \frac{\Delta_\epsilon f(x)}{\epsilon} \leqq \frac{\Delta_\epsilon f(y)}{\epsilon}.$$

Let $\delta \to 0+$; (6) shows that

$$\text{(7)} \qquad \frac{\Delta_\epsilon f(z - \epsilon)}{\epsilon} \leqq f'_-(x) \leqq f'_+(x) \leqq \frac{\Delta_\epsilon f(y)}{\epsilon}.$$

Hence $f'_+(x)$ and $f'_-(x)$ are finite. Now let $\epsilon \to 0+$; (7) shows that

$$f'_-(z) \leqq f'_-(x) \leqq f'_+(x) \leqq f'_+(y) \qquad (a < z < x < y < b).$$

Hence $f'_+(x)$ and $f'_-(x)$ are non-decreasing.

LEMMA 5. *$f(x)$ approaches a limit or becomes positively infinite as $x \to a$ and as $x \to b$.*

By Lemma 2, $\Delta_\delta f(x)$ is a non-decreasing function of x in $(a, b - \delta)$. Hence $\lim_{x \to a+} \Delta_\delta f(x)$ exists unless $\Delta_\delta f(x) \to -\infty$ $(x \to a+)$. Thus

$$\lim_{x \to a+} \Delta_\delta f(x) = \lim_{x \to a+} [f(x + \delta) - f(x)] = f(\delta) - f(a+) < +\infty,$$

so that $f(a+)$ exists unless $f(x) \to +\infty$ as $x \to a$.

Similarly, as $x \to b - \delta -$, $\Delta_\delta f(x) = f(x + \delta) - f(x)$ approaches a limit or becomes positively infinite; $f(x) \to f(b - \delta)$; hence $f(b-)$ exists unless $f(x) \to +\infty$ as $x \to b$.

LEMMA 6. *If $f(a+) < \infty$ and $f(a) = f(a+)$, $f'_+(a)$ exists (finite or $+\infty$). If $f(b-) < \infty$ and $f(b) = f(b-)$, $f'_-(b)$ exists (finite or $-\infty$).*

By Lemma 3, $h^{-1}\Delta_h f(x)$ is a non-decreasing function of h in $0 < h < b - x$. Since $f(a) = f(a+) < \infty$, $h^{-1}\Delta_h f(a)$ is the limit of a sequence of non-decreasing functions, and consequently is non-decreasing; the first conclusion follows.

Similarly, $h^{-1}\Delta_h f(b - h)$ is a non-increasing function of h, and the second conclusion follows.

Common hypothesis of Lemmas 7–14: $f(x)$ satisfies H_k in (a, b), with $k > 2$.

LEMMA 7. *If $a < x < b$, $h^{-k+1}\Delta_h^{k-1} f(x)$ is a non-decreasing function of h in $0 < h < (b - x)/(k - 1)$.*

By Lemma 1 with $\delta_1 = \delta_2 = \delta$, $\delta_3 = \cdots = \delta_k = h$,

$$\Delta_\delta^2 \Delta_h^{k-2} f(x) \geqq 0 \qquad (x > a,\ x + 2\delta + (k-2)h < b).$$

Thus if h is fixed, $\Delta_h^{k-2} f(x)$ satisfies H_2 in $a < x < b - (k-2)h$. By Lemma 3, if x is fixed,

(8) $\dfrac{1}{\delta} \Delta_\delta \Delta_h^{k-2} f(x)$ is a non-decreasing function of δ in $0 < \delta < b - (k-2)h - x$.

By Lemma 1 with $\delta_1 = \delta_2 = \cdots = \delta_{k-1} = \delta$, $\delta_k = h$,

$$\Delta_h^{k-1} \Delta_\delta f(x) \geqq 0 \qquad (x > a,\ x + \delta + (k-1)h < b).$$

Thus if δ is fixed

$$\Delta_\delta f(x) \text{ satisfies } H_{k-1} \text{ in } a < x < b - \delta. \tag{9}$$

We now assume Lemma 7 for $k - 1$ and prove it for k. If x is fixed, $x + (k-1)\epsilon < b$, and $0 < \delta < \epsilon$,

$$\frac{\Delta_\epsilon^{k-1} f(x)}{\epsilon^{k-1}} = \frac{1}{\epsilon^{k-2}} \frac{\Delta_\epsilon \Delta_\epsilon^{k-2} f(x)}{\epsilon} \geqq \frac{1}{\epsilon^{k-2}} \frac{\Delta_\delta \Delta_\epsilon^{k-2} f(x)}{\delta} = \frac{1}{\delta} \frac{\Delta_\epsilon^{k-2} \Delta_\delta f(x)}{\epsilon^{k-2}} \tag{10}$$

by (8). By (9) and the induction hypothesis,

$$\frac{1}{\delta} \frac{\Delta_\epsilon^{k-2} \Delta_\delta f(x)}{\epsilon^{k-2}} \geqq \frac{1}{\delta} \frac{\Delta_\delta^{k-2} \Delta_\delta f(x)}{\delta^{k-2}} = \frac{\Delta_\delta^{k-1} f(x)}{\delta^{k-1}}. \tag{11}$$

Combining (9) and (10), we have the conclusion of Lemma 7 for k. Since Lemma 7 is true for $k = 2$ (Lemma 3), it is true for all $k > 2$.

Lemma 8. *There is a point c, $a \leqq c \leqq b$, such that $f(x)$ satisfies H_{k-1} in $c < x < b$ (if $c < b$), and $-f(x)$ satisfies H_{k-1} in $a < x < c$ (if $c > a$).*

Divide the points x of $a < x < b$ into classes A and B as follows:

$$x \in A \text{ if } \Delta_\delta^{k-1} f(x) \geqq 0 \text{ whenever } 0 < (k-1)\delta < b - x,$$

$$x \in B \text{ if } \Delta_\delta^{k-1} f(x) < 0 \text{ for some } \delta,\ 0 < (k-1)\delta < b - x.$$

A and B are mutually exclusive and exhaust (a, b); either may be empty. If neither is empty, and $z \in A$, $y \in B$, then $y < z$. For, since $y \in B$, there is an ϵ such that $y < y + (k-1)\epsilon < b$ and $\Delta_\epsilon^{k-1} f(y) < 0$. By Lemma 2, if $z < y$,

$$\Delta_\epsilon^{k-1} f(z) \leqq \Delta_\epsilon^{k-1} f(y) < 0,$$

and this contradicts $z \in A$. Hence $y < z$.

From these properties of A and B it follows that there is a real number c such that $a \leqq c \leqq b$, $x \leqq c$ if $x \in B$, $x \geqq c$ if $x \in A$. If $c < b$, $f(x)$ satisfies H_{k-1} in $c < x < b$, by the definition of A.

If $c > a$ and $a < x < c$, $\Delta_\epsilon^{k-1} f(x) < 0$ for some ϵ, $x < x + (k-1)\epsilon < b - x$. By Lemmas 7 and 2, $\Delta_\delta^{k-1} f(y) < 0$ if $a < y < x$ and $0 < \delta < \epsilon$. Thus for each x in $a < x < c$ there is a positive $\epsilon(x)$ such that $\Delta_\delta^{k-1} f(x) < 0$ for $0 < \delta <$

$\epsilon(x)$, and $\epsilon(x)$ is a non-increasing function of x. Now let $a < x < c$ and $0 < (k-1)\delta < c - x$, and choose y, $x + (k-1)\delta < y < c$. Take an integer n so large that $\delta/n < \epsilon(y)$. Then

$$\Delta_\delta^{k-1} f(x) = \sum_{i=0}^{n-1} \Delta_{\delta/n}^{k-1} f(x + i\delta/n) < 0,$$

since $\delta/n < \epsilon(y) \leqq \epsilon(x + i\delta/n)$ $(i = 0, 1, \cdots, n-1)$. Thus $-f(x)$ satisfies H_{k-1} in (a, c).

LEMMA 9. *There are points $x_0, x_1, \cdots, x_p$, $a = x_0 < x_1 < \cdots < x_p = b$, $1 \leqq p \leqq 2^{k-1}$, such that in each interval $x_i < x < x_{i+1}$ either $f(x)$ or $-f(x)$ satisfies H_2.*

This follows from Lemma 8 by induction on k.

LEMMA 10. *$f'_+(x)$ exists and is finite in $a < x < b$.*

By Lemmas 9, 4, and 6, $f'_+(x)$ exists in (a, b), and is finite except perhaps at the points x_j $(j = 1, 2, \cdots, p)$ of Lemma 9. Suppose that for some j $(1 \leqq j \leqq p-1)$ $f'_+(x_j) = +\infty$. If δ satisfies $x_{j-1} < x_j - k\delta < x_j + \delta < x_{j+1}$,

$$\Delta_\delta^{k-1} f'_+[x_j - (k-2)\delta] = -\infty;$$

$$\lim_{h \to 0+} \frac{1}{h} \Delta_\delta^{k-1} \Delta_h f[x_j - (k-2)\delta] = -\infty;$$

$$\Delta_h \Delta_\delta^{k-1} f[x_j - (k-2)\delta] < 0$$

for all small positive h. This contradicts Lemma 1 (with $\delta_1 = h$, $\delta_2 = \delta_k = \cdots = \delta$).

If $f'_+(x_j) = -\infty$,

$$\Delta_\delta^{k-1} f'_+[x_j - (k-1)\delta] = -\infty \qquad (x_{j-1} < x_j - (k-1)\delta < x_j),$$

and consequently

$$\Delta_h \Delta_\delta^{k-1} f[x_j - (k-1)\delta] < 0$$

for small positive h, and again Lemma 1 is contradicted.

LEMMA 11. *$f'(x)$ exists and is finite in $a < x < b$.*

For fixed x, let $a < a + kh < x < b - 2h$. Let p be either $k-1$ or $k-2$. Then

$$\Delta_h^k f(x - ph) = \sum_{i=0}^{k} (-1)^{k-i} \binom{k}{i} f[x + (i-p)h] \geqq 0.$$

Since

$$\sum_{i=0}^{k} (-1)^{k-i} \binom{k}{i} = (-1)^k (1-1)^k = 0,$$

$$\sum_{\substack{i=0 \\ i \neq p}}^{k} (-1)^{k-i} \binom{k}{i} \frac{f[x + (i-p)h] - f(x)}{(i-p)h} (i-p) \geqq 0.$$

Letting $h \to 0+$, we deduce

$$A_p f'_-(x) + B_p f'_+(x) \geqq 0, \tag{12}$$

where

$$A_p = \sum_{i=0}^{p-1} (-1)^{k-i} \binom{k}{i} (i - p), \qquad B_p = \sum_{i=p+1}^{k} (-1)^{k-i} \binom{k}{i} (i - p),$$

$$A_p + B_p = \sum_{i=0}^{k} (-1)^{k-i} \binom{k}{i} (i - p)$$

$$= k \sum_{i=1}^{k} (-1)^{k-i} \binom{k-1}{i} + p \sum_{i=0}^{k} (-1)^{k-i} \binom{k}{i} = 0.$$

If $p = k - 1$, $B_p = 1$ and $A_p = -1$; from (12),

$$f'_+(x) \geqq f'_-(x).$$

If $p = k - 2$, $B_p = -k + 2 < 0$ (since $k > 2$), and $A_p = k - 2 > 0$; from (12),

$$f'_-(x) \geqq f'_+(x).$$

Hence $f'_-(x) = f'_+(x)$. That is, $f'(x)$ exists; it is finite by Lemma 10.

LEMMA 12. *$\Delta_h^{k-1} f'(x) \geqq 0$ if $a < x < x + (k - 1)h < b$.*

For,

$$\Delta_h^{k-1} f'(x) = \lim_{\delta \to 0+} \frac{\Delta_\delta \Delta_h^{k-1} f(x)}{\delta} \geqq 0$$

by Lemma 1.

LEMMA 13. *$f'(x)$ is continuous in $a < x < b$.*

Since $f'(x)$ is monotonic, $f'(x+)$ and $f'(x-)$ exist. Let x be fixed, let $a < x - kh < x < x + 2h < b$, and let p be $k - 1$ or $k - 2$. By Lemma 12,

$$\Delta_h^{k-1} f'[x - (p - \tfrac{1}{2})h] = \sum_{i=0}^{k-1} (-1)^{k-i-1} \binom{k-1}{i} f'[x + (i - p + \tfrac{1}{2})h] \geqq 0.$$

Let $h \to 0+$. Then

$$A_p f'(x-) - A_p f'(x+) \geqq 0, \tag{13}$$

where

$$A_p = \sum_{i=0}^{p-1} (-1)^{k-i-1} \binom{k-1}{i} = - \sum_{i=p}^{k-1} (-1)^{k-i-1} \binom{k-1}{i}.$$

If $p = k - 1$, $A_p = -1$; if $p = k - 2$, $A_p = k - 2 > 0$. Hence

$$f'(x-) = f'(x+)$$

follows from (13).

Moreover,

$$\Delta_h^{k-1} f'(x - ph) \geqq 0,$$

and hence

$$-B_p f'(x+) + B_p f'(x) \geqq 0,$$

where

$$B_p = (-1)^{k-p} \binom{k-1}{p} = -\sum_{\substack{i=0 \\ i \neq p}}^{k-1} (-1)^{k-i-1} \binom{k-1}{i}.$$

For $p = k - 1$ and $p = k - 2$, B_p is respectively $+1$ and $-(k - 1)$; hence $f'(x) = f'(x+) = f'(x-)$; that is, $f'(x)$ is continuous.

LEMMA 14. *$f'(x)$ satisfies H_{k-1} in (a, b).*

The lemma follows from Lemmas 12 and 13.

Proof of the theorem. We prove the theorem with the aid of Lemmas 14 and 4, using induction on k.

DUKE UNIVERSITY AND HARVARD UNIVERSITY.

SYMMETRIC TRANSFORMATIONS IN HILBERT SPACE

By J. W. Calkin

At various points in the further development of the abstract theory of boundary conditions [1],[1] and in the applications of this theory to differential equations, certain elementary questions concerning symmetric transformations[2] in Hilbert space have arisen which have escaped formal notice in the previous literature. It is the purpose of this note to dispose of these matters.

The questions with which we are concerned refer to a closed linear symmetric transformation H with the property that $(H - \lambda I)^{-1}$ exists and is bounded for some real λ. Such a symmetric transformation always possesses a self-adjoint extension, and here we show that the deficiency-index of H is (n, n), where n is the dimension number of the manifold of solutions of the equation $H^*f - \lambda f = 0$. In addition, we show that under the condition stated, H has a self-adjoint extension S with λ in its resolvent set. Finally, assuming one such extension to be known, we obtain a characterization of all maximal symmetric extensions of H in terms of maximal symmetric transformations in the manifold of zeros of $H^* - \lambda I$ and its subspaces.

We use throughout the paper the notation and terminology of the treatise of Stone [6], except for minor modifications; in particular, we denote the domain, range, and graph of a transformation T in Hilbert space $\mathfrak{H}$ by $\mathfrak{D}(T)$, $\mathfrak{R}(T)$, and $\mathfrak{G}(T)$, respectively.

Before proceeding, we note a simple theorem which permits us, without affecting the generality of the results, to reduce our problem to the case that the number λ appearing in the condition described above is zero.

Theorem 1. *Let H be a closed linear symmetric transformation in Hilbert space $\mathfrak{H}$. Then S is a symmetric extension of H if and only if $S - \lambda I$ is a symmetric extension of $H - \lambda I$ ($\mathfrak{J}(\lambda) = 0$). Moreover S and $S - \lambda I$ have the same deficiency-index.*

The first assertion is obvious, and the second is an immediate consequence of elementary facts in the general theory of symmetric transformations.[3]

We consider then a closed linear symmetric transformation H such that H^{-1} is bounded and proceed to construct a self-adjoint extension S of H, also with bounded inverse. In order to effect this construction we first note that since H^{-1} is bounded, $\mathfrak{R}(H)$ is closed. Thus H^{-1} can be regarded as a linear trans-

Received January 27, 1939; in revised form, September 3, 1940.

[1] Numbers in brackets designate the references listed at the end of the paper.

[2] For the general theory of symmetric transformations see [4], especially pp. 80–91, or [6], Chapter 9.

[3] [6], Theorem 9.8.

formation from the space $\mathfrak{R}(H)$ to $\mathfrak{H}$ in the sense of Murray [3] and possesses in this sense an adjoint T with domain in $\mathfrak{H}$ and range in $\mathfrak{R}(H)$. Moreover, the boundedness of H^{-1} implies that T is bounded and has domain identically $\mathfrak{H}$, while the fact that H^{-1} has range dense in $\mathfrak{H}$ implies that T^{-1} exists.[4] But T^{-1} can also be regarded as a linear transformation in $\mathfrak{H}$ and from this point of view we have clearly $T^{-1} \subseteq H^*$.

Now let f be an element of $\mathfrak{D}(H)$ which is also in $\mathfrak{D}(T^{-1})$. The latter assumption implies that f is in $\mathfrak{R}(H)$ and since

$$(g, f) - (H^{-1}g, Hf) = 0$$

for all g in $\mathfrak{R}(H)$, it follows that Hf is in $\mathfrak{D}(T)$ and that $f = THf$. Thus, where H and T^{-1} are both defined they are equal, and the linear transformation S with domain $\mathfrak{D}(H) + \mathfrak{D}(T^{-1})$ which is equal to H on $\mathfrak{D}(H)$ and to T^{-1} on $\mathfrak{D}(T^{-1})$ is single-valued. Moreover, $H \subseteq S$ and $\mathfrak{R}(S) = \mathfrak{H}$.

We shall now show that S is symmetric. That its domain is dense in $\mathfrak{H}$ follows from the relation $H \subseteq S$; accordingly we have to prove that

$$(f, Sg) - (Sf, g) = 0 \tag{1}$$

for all f and g in $\mathfrak{D}(S)$. To do this we set $Sf = u_1 + v_1$, $Sg = u_2 + v_2$, where u_1 and u_2 are in $\mathfrak{R}(H)$, v_1 and v_2 in $\mathfrak{H} \ominus \mathfrak{R}(H)$. Then, by virtue of our definition of S, we have

$$f = H^{-1}u_1 + Tv_1, \qquad g = H^{-1}u_2 + Tv_2,$$

and the condition (1) becomes

$$(H^{-1}u_1 + Tv_1, u_2 + v_2) - (u_1 + v_1, H^{-1}u_2 + Tv_2) = 0,$$

or

$$\begin{aligned} &[(H^{-1}u_1, u_2) - (u_1, H^{-1}u_2)] + [(H^{-1}u_1, v_2) - (u_1, Tv_2)] \\ &\qquad + [(Tv_1, u_2) - (v_1, H^{-1}u_2)] + (Tv_1, v_2) - (v_1, Tv_2) = 0 \end{aligned} \tag{2}$$

But in (2) the first bracketed expression vanishes because of the symmetry of H, while the other two vanish by reason of the adjoint relationship between H^{-1} and T. Finally, the last two inner products vanish because of the fact that v_1 and v_2 are in $\mathfrak{H} \ominus \mathfrak{R}(H)$, while $\mathfrak{R}(T) \subseteq \mathfrak{R}(H)$. Thus (2) is true and therefore (1) is also. Hence S is symmetric and this fact coupled with the previous result, $\mathfrak{R}(S) = \mathfrak{H}$, allows us to conclude that S is self-adjoint[5] and, at the same time, that its resolvent set contains the origin.

Hence, taking account of Theorem 1, we have

THEOREM 2. *Let H be a closed linear symmetric transformation in $\mathfrak{H}$ such that $(H - \lambda I)^{-1}$ exists and is bounded for some real λ. Then there exists a self-adjoint extension of H with λ in its resolvent set.*

[4] [3], Theorem II.
[5] [6], Theorem 2.19.

Returning to the assumption that H^{-1} exists and is bounded, we observe that Theorem 2 implies that the deficiency-index of H is of the form (n, n), and recall that by definition n is the dimension number of the manifold $\mathfrak{D}^+$ of solutions of the equation $H^*f = if$.[6] We then invoke a result of Teichmüller [7][7] to conclude that $\mathfrak{B}(S) \ominus \mathfrak{B}(H)$ consists of all elements of $\mathfrak{H} \oplus \mathfrak{H}$ which have the form $\{f^+ - Vf^+, i(f^+ + Vf^+)\}$, where f^+ is in $\mathfrak{D}^+$ and V is the (unitary) Cayley transform of S. Hence it follows that $\mathfrak{B}(S) \ominus \mathfrak{B}(H)$ has the dimension number n and that $\mathfrak{B}(S^{-1}) \ominus \mathfrak{B}(H^{-1})$ does also. But the latter manifold clearly has the same dimension number as $\mathfrak{R}(S) \ominus \mathfrak{R}(H) = \mathfrak{H} \ominus \mathfrak{R}(H)$ and this is precisely the manifold of zeros of H^*.[8] Accordingly, again generalizing our conclusion by means of Theorem 1, we obtain

THEOREM 3. *If the hypothesis of Theorem 2 is satisfied, the deficiency-index of H is (n, n), where n is the dimension number of the manifold of zeros of $H^* - \lambda I$.*

We proceed now to a constructive characterization of the maximal symmetric extensions of a closed linear symmetric transformation H with bounded inverse. Theorem 2 provides us with an extension S of H with the origin in its resolvent set; alternatively, any other extension S with that property may be taken as basic in the construction. To begin, we consider an arbitrary maximal symmetric extension R of H and denote by $\mathfrak{M}$ the manifold of zeros of R^*, observing that the maximality of R implies $Rf = 0$ whenever $R^*f = 0$. By R_0, we designate the transformation induced in $\mathfrak{H} \ominus \mathfrak{M}$ by R; evidently R_0 is maximal symmetric and hence R_0^{-1}, which clearly exists, is maximal also, and self-adjoint if and only if R is.

Now let E be the projection with range $\mathfrak{H} \ominus \mathfrak{M}$, and let S_0 be the transformation induced in $\mathfrak{H} \ominus \mathfrak{M}$ by ESE. Clearly, S_0 is self-adjoint in $\mathfrak{H} \ominus \mathfrak{M}$, and since its range is $\mathfrak{H} \ominus \mathfrak{M}$, S_0^{-1} exists and is bounded. Moreover, the boundedness of S_0^{-1} implies $(R_0^{-1} - S_0^{-1})^* \equiv (R_0^{-1})^* - S_0^{-1}$, and therefore $K_R \equiv R_0^{-1} - S_0^{-1}$ is self-adjoint if and only if R is. To obtain a more precise result let us consider the deficiency-index (n, m) of R. Then (n, m) is evidently the deficiency-index of R_0 also, and by straightforward algebraic calculation (m, n) is the deficiency-index of R_0^{-1}. Now let A be an arbitrary reduction operator for $(R_0^{-1})^*$, W the unitary operator in the range-space of A associated with A.[9] Then A is also, in effect, a reduction operator for $(R_0^{-1} - S_0^{-1})^*$, by Theorem 1.3 of the paper just referred to. But by Theorem 3.7 of the same paper, the deficiency-indices of both R_0^{-1} and $R_0^{-1} - S_0^{-1}$ are completely determined by W; that is to say, K_R and R_0^{-1} have the same deficiency-index. Finally, we note that $K_R = 0$ in $\mathfrak{R}(H)$; thus its character is completely determined by its behavior in $\mathfrak{H} \ominus (\mathfrak{R}(H) \oplus \mathfrak{M})$.

Now let $\mathfrak{M}$ be an arbitrary closed linear manifold in $\mathfrak{H} \ominus \mathfrak{R}(H)$, the manifold

[6] [6], Theorem 9.3 and Definition 9.1.

[7] See also [1], Theorems 2.8 and 4.2.

[8] By [5], Satz 7, for example.

[9] For the definition of these concepts and their elementary analysis, see [1], Chapter 1.

of zeros of H^*, and let K be an arbitrary maximal symmetric transformation in $\mathfrak{H} \ominus \mathfrak{M}$ which has $\mathfrak{N}(H)$ in its manifold of zeros. Let E be the projection with range $\mathfrak{H} \ominus \mathfrak{M}$ and let S_0 be the transformation induced in $\mathfrak{H} \ominus \mathfrak{M}$ by ESE, where S is a self-adjoint extension of H with bounded inverse. Let $L = S_0^{-1} + K$. Then $Lf = 0$ implies that Kf is in the domain of S_0, but since Kf is also in $\mathfrak{H} \ominus \mathfrak{N}(H)$, we have $H^*Kf = 0$ which implies $S_0 f = 0$, and this in turn implies $f = 0$. Thus L^{-1} exists. But, on $\mathfrak{N}(H)$, $L = S_0^{-1} = EH^{-1}E$, and hence L^{-1} is an extension of EHE in $\mathfrak{H} \ominus \mathfrak{M}$. Moreover, by a repetition of the argument of the preceding paragraph, the deficiency-index of L^{-1} is (n, m), where (m, n) is the deficiency-index of K. Finally, let R be the transformation which is equal to L^{-1} in $\mathfrak{H} \ominus \mathfrak{M}$ and takes every element of $\mathfrak{M}$ into zero. Then R has the same deficiency-index as L^{-1} and thus is a maximal symmetric extension of H; moreover, if K_R is the transformation in $\mathfrak{H} \ominus \mathfrak{M}$ associated with R above, then K_R is identical with K. Thus, summing up in a formal statement, we have

THEOREM 4. *Let H be a closed symmetric transformation in $\mathfrak{H}$ such that H^{-1} exists and is bounded, and let S be a self-adjoint extension of H with the same property. Let $\mathfrak{M}$ be an arbitrary closed linear manifold in $\mathfrak{H} \ominus \mathfrak{N}(H)$, the manifold of zeros of H, E the projection operator of $\mathfrak{H} \ominus \mathfrak{M}$. Then the class $\mathfrak{S}_{\mathfrak{M}}$ of all maximal symmetric extensions R of H which have $\mathfrak{M}$ as characteristic manifolds corresponding to the characteristic value zero is in one-to-one correspondence with the class $\mathfrak{K}_{\mathfrak{M}}$ of all maximal symmetric transformations K in $\mathfrak{H} \ominus \mathfrak{M}$ which are equal to zero on $\mathfrak{N}(H)$. If R and K are corresponding members of $\mathfrak{S}_{\mathfrak{M}}$ and $\mathfrak{K}_{\mathfrak{M}}$, respectively, and R_0 is the operator induced in $\mathfrak{H} \ominus \mathfrak{M}$ by R, S_0 the operator induced in $\mathfrak{H} \ominus \mathfrak{M}$ by ESE, then*

$$K = R_0^{-1} - S_0^{-1}.$$

The deficiency-index of R is $(0, n)$ or $(m, 0)$ according as the deficiency-index of K is $(n, 0)$ or $(0, m)$. In particular, R is self-adjoint if and only if K is, and has a bounded inverse if and only if $\mathfrak{M} = [0]$ and K is bounded.

While the direct proof of Theorem 4 which we have given is more straightforward and no more complicated, it is nevertheless interesting to note that another proof can be based on the general analysis of [1]. To obtain this, one begins by considering the operator A with domain $\mathfrak{D}(H^*)$ which takes $\{f, H^*f\}$ into $\{-f + S^{-1}H^*f, -EH^*f\}$, where E is the projection operator of the manifold $\mathfrak{U}$ of zeros of H^*. It can then be established that $\mathfrak{R}(A) = \mathfrak{U} \oplus \mathfrak{U}$ and that A is in fact a bounded reduction operator for H^* with range-space $\mathfrak{U} \oplus \mathfrak{U}$. The isometric transformation W associated with A by Definition 1.1 of [1] is simply the transformation which takes $\{u, v\}$ into $\{v, -u\}$. Thus Theorem 4.1 of [1] provides a characterization of the maximal symmetric extensions of H in terms of W-symmetric manifolds in $\mathfrak{U} \oplus \mathfrak{U}$ and the analysis of these manifolds provided by Theorem 2.10 of the same paper yields at once Theorem 4 above.

It is worth while also to point out in connection with Theorem 4 that a slight elaboration of the construction on which it is based yields an analogous characterization of all closed linear symmetric extensions of H. This becomes even more clear on the basis of the remarks of the preceding paragraph.

In conclusion, we call attention to certain implications of the preceding analysis.

First, it may be noted that the familiar theorem that every definite closed linear symmetric transformation possesses a self-adjoint extension [4, 6, 2] is an immediate corollary of Theorem 2; in fact, von Neumann's proof of the former theorem makes essential use only of the hypothesis of our Theorem 2 and suggests an alternative proof of our Theorem 3.

Secondly, it is worth observing that the hypothesis of Theorem 2 is equivalent to the somewhat more complicated hypothesis of Theorem 9.21 of the book of Stone; in fact, although quite different in motivation, Theorem 2 lies very close to the latter theorem. The equivalence of the hypotheses, however, does not appear to be obvious except on the basis of the conclusions of the two theorems.

Finally, we wish to point out that, except for those portions of the analysis which are concerned with establishing the relation between the constructions of the present paper and the von Neumann theory of Cayley transforms, no essential use has been made of the fact that we are dealing with complex Hilbert space. Accordingly, precise analogues of the preceding constructions are possible for both the real and quaternionic cases.

References

1. J. W. Calkin, *Abstract symmetric boundary conditions*, Transactions of the American Mathematical Society, vol. 45(1939), pp. 369–442.
2. K. Friedrichs, *Spektraltheorie halbbeschränkter Operatoren*, Mathematische Annalen, vol. 109(1934), pp. 465–487.
3. F. J. Murray, *Linear transformations between Hilbert spaces and the application of this theory to linear partial differential equations*, Transactions of the American Mathematical Society, vol. 37(1935), pp. 301–338.
4. J. von Neumann, *Allgemeine Eigenwerttheorie Hermitescher Funktionaloperatoren*, Mathematische Annalen, vol. 102(1929), pp. 49–131.
5. J. von Neumann, *Über adjungtierte Funktionaloperatoren*, Annals of Mathematics, (2), vol. 33(1932), pp. 294–310.
6. M. H. Stone, *Linear Transformations in Hilbert Space and Their Applications to Analysis*, New York, 1932.
7. O. Teichmüller, *Operatoren in Wachsen Raum*, Journal für die reine und angewandte Mathematik, vol. 174(1935), pp. 73–124.

Illinois Institute of Technology.

End
Kay